

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»**

*Факультет інформатики та обчислювальної техніки  
Кафедра автоматизованих систем обробки інформації та управління*

"На правах рукопису"

УДК 519.854.2

До захисту допущено

В.о. завідувача кафедри

\_\_\_\_\_ Олександр ПАВЛОВ\_

“ \_\_\_\_\_ ” \_\_\_\_\_ 20 20 р.

**МАГІСТЕРСЬКА ДИСЕРТАЦІЯ**

**на здобуття ступеня магістра**

**за освітньо-професійною програмою**

***«Інформаційні управляючі системи та технології»***

**зі спеціальності 126 *«Інформаційні системи та технології»***

**на тему:**

**«Аналіз моделей ранжування вершин у графах мереж різного  
призначення»**

Виконала:

Студентка VI курсу, групи ІС-92мп

Вихляєва Аліса Олегівна \_\_\_\_\_

Керівник:

доц., к.т.н., доц.

Попенко Володимир Дмитрович \_\_\_\_\_

Консультант:

професор, д.т.н., доцент,

Жаріков Едуард В'ячеславович \_\_\_\_\_

Рецензент:

доц., к.т.н., доц.

Репнікова Наталія Борисівна \_\_\_\_\_

Засвідчую, що у цій магістерській  
дисертації немає запозичень з праць  
інших авторів без відповідних посилань.  
Студентка \_\_\_\_\_

Київ – 2020 року

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»**

*Факультет інформатики та обчислювальної техніки  
Кафедра автоматизованих систем обробки інформації та управління*

Рівень вищої освіти – *другий (магістерський)*

Спеціальність – *126 «Інформаційні системи та технології»*

Освітньо-професійна програма *«Інформаційні управляючі системи та технології»*

В.о.з авідувача кафедри

\_\_\_\_\_ Олександр ПАВЛОВ

«\_\_» \_\_\_\_\_ 2020 р.

**ЗАВДАННЯ  
на магістерську дисертацію студенту**

**Вихляєва Аліса Олегівна**

1. Тема дисертації «Аналіз моделей ранжування вершин у графах мереж різного призначення», науковий керівник дисертації Попенко Володимир Дмитрович к.т.н., доцент, затверджені наказом по університету від «26» жовтня 2020 р. № 3132-с

2. Строк подання студентом дисертації “   2   ”   12     20     20   р.

3. Об’єкт дослідження - процес аналізу мереж різного призначення

4. Перелік завдань, які потрібно розробити:

*Проаналізувати існуючі методи дослідження соціальних мереж, провести аналіз відомих робіт з розв’язання поставленої в рамках роботи задачі, дослідити центральність вершин графа в моделі електричних ланцюгів з використанням кооперативної теорії ігор, провести порівняльний аналіз методів ранжування вершин у графах, виконати програмну реалізацію методів ранжування вершин у графах, дослідити ефективність алгоритмів та методів при різних вхідних даних шляхом проведення обчислювальних експериментів, проаналізувати дані експериментів та порівняти з вже існуючими рішеннями.*

## 5. Орієнтовний перелік графічного (ілюстративного) матеріалу

1. Схема структурна контекстної моделі системи. 2. Діаграма діяльності роботи програми. 3. Діаграма послідовності роботи розробленого застосунку. 4. Схема структурна класів програмного забезпечення. 5. Схема структурна варіантів використання. 6. Екранні форми. 7. Результати експериментів.

## 6. Орієнтовний перелік публікацій

Дві публікації: одна стаття у фаховому та одні тези доповіді на науково-практичній конференції

## 7. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

8. Дата видачі завдання “ 1 ” вересня 20 20 р.

## Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Строк виконання етапів магістерської дисертації	Примітка
1	Систематизація результатів огляду літератури	15.09	Отримати підпис керівника
2	Порівняльний аналіз існуючих методів розв'язання задачі	26.09	Отримати підпис керівника
3	Постановка та формалізація математичної моделі задачі	10.10	Отримати підпис керівника
4	Модифікація існуючих методів розв'язання задачі	20.1	Отримати підпис керівника
5	Розробка інформаційного та програмного забезпечення	01.1	Отримати підпис керівника
7	Проведення експериментальних досліджень розроблених алгоритмів	09.11	Отримати підпис керівника
8	Оформлення документації	17.11	Отримати підпис керівника
9	Подання роботи на попередній захист	20.11	Отримати підпис керівника
10	Подання роботи на основний захист	02.12	Отримати підпис керівника

Студент

Аліса ВИХЛЯЄВА

Науковий керівник

Володимир ПОПЕНКО

## РЕФЕРАТ

Магістерська дисертація: 82 с., 5 рис., 27 табл., 74 джерел, 1 додаток.

**Актуальність.** *Актуальність* даної теми обумовлена тим, що теорія соціальних мереж дає відповідь на питання, яке займало соціальну філософію ще з часів Платона, а саме проблему соціального порядку: як автономні індивіди можуть поєднуватися для створення стійких, функціонуючих суспільств. Теорія мережі також дає пояснення для безлічі соціальних явищ: від особистої творчості до корпоративної прибутковості. З аналізом соціальних мереж пов'язані феномени соціалізації персональних даних, стали публічно доступними факти біографії, переписки, особисті статки, щоденники, фото-, відео-, аудіоматеріали, новини світу, державні таємниці, нотатки про подорожі та інше. Таким чином, у соціальних мережах є унікальні дані про особисту історію та інтереси реальних людей. Аналіз соціальних мереж часто використовується в таких сферах, як обстеження індивідуальних та соціальних груп, кластеризація (визначення відносин), електронна комерція та інтернетреклама (профілювання клієнтів, аналіз тенденцій, персоналізована реклама та подання пропозицій), аналіз фізичних структур (транспорт, встановлення, інфраструктура) та аналіз великих наборів даних (спостереження за ЗМІ, аналіз академічних публікацій, генетичні дослідження). Це відкриває безпрецедентні можливості для вирішення дослідницьких і бізнес-завдань, а також створення допоміжних сервісів і застосунків для користувачів соціальних мереж. Крім того, це пояснює підвищений інтерес до збору та аналізу соціальних даних зі сторони компаній та дослідницьких центрів.

**Зв'язок роботи з науковими програмами, планами, темами.** Робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Аналіз моделей ранжування вершин у графах мереж різного призначення» (№ ДР 0117U0009100).

**Мета дослідження** – побудова і дослідження властивостей математичних моделей ранжування вершин в графах мереж різного призначення.

Для досягнення мети необхідно виконати наступні **завдання**:

- проаналізувати існуючі методи дослідження соціальних мереж;
- провести аналіз відомих робіт з розв'язання поставленої в рамках роботи задачі;
- дослідити центральність вершин графа в моделі електричних ланцюгів з використанням кооперативної теорії ігор;
- провести порівняльний аналіз методів ранжування вершин у графах;
- виконати програмну реалізацію методів ранжування вершин у графах;
- дослідити ефективність алгоритмів та методів при різних вхідних даних шляхом проведення обчислювальних експериментів;
- проаналізувати дані експериментів та порівняти з вже існуючими рішеннями.

**Об'єкт дослідження** – процес аналізу мереж різного призначення.

**Предмет дослідження** – моделі та методи аналізу мереж різного призначення.

**Наукова новизна одержаних результатів** полягає у дослідженні математичних моделей ранжування вершин графа, а також у впровадженні системи, яка проводить порівняльний аналіз цих моделей.

**Публікації.** Вихляєва А., Попенко В.Д. Аналіз соціальних мереж // *Perspectives of world science and education. Abstracts of the 9th International scientific and practical conference. CPN Publishing Group. Osaka, Japan. 2020. Pp. 375-391.*

СОЦІАЛЬНА МЕРЕЖА, ВЕРШИНА, РАНЖУВАННЯ, ЕЛЕКТРИЧНА ЦЕНТРАЛЬНІСТЬ, ЦЕНТРАЛЬНІСТЬ, PAGERANK.

## ABSTRACT

Master's dissertation: 82 p., 5 figs., 27 tables, 74 sources, 1 appendix. **Relevance.** The relevance of this topic is due to the fact that the theory of social networks provides an answer to the question that has occupied social philosophy since the time of Plato, namely the problem of social order: how autonomous individuals can combine to create stable, functioning societies. Network theory also provides an explanation for many social phenomena: from personal creativity to corporate profitability. The analysis of social networks is associated with the phenomena of socialization of personal data: biographies, correspondence, personal belongings, diaries, photos, videos, audio, world news, state secrets, travel notes and more have become publicly available. Thus, social networks have unique information about the personal history and interests of real people. Analysis of social networks is often used in such areas as surveys of individual and social groups, clustering (definition of relationships), e-commerce and Internet advertising (customer profiling, trend analysis, personalized advertising and submission of proposals), analysis of physical structures (transport, installation, infrastructure) and analysis of large data sets (media observation, analysis of academic publications, genetic research). This opens up unprecedented opportunities for research and business solutions, as well as the creation of support services and applications for social network users. In addition, this explains the increased interest in the collection and analysis of social data by companies and research centers.

**Connection of work with scientific programs, plans, themes.** The work was performed at the Department of Automated Information Processing and Control Systems of the National Technical University of Ukraine "Kyiv Polytechnic Institute. Igor Sikorsky "within the topic" Analysis of vertex ranking models in graphs of networks for different purposes "(№ DR 0117U0009100).

**The purpose of the study** is to construct and study the properties of mathematical models of vertex ranking in graphs of a network for various purposes.

To achieve this goal you must perform the following tasks: - analyze existing research methods of social networks;

- to analyze the known work to solve the problem set in the work;
- to investigate the centrality of graph vertices in the model of electric circuits using cooperative game theory
- to conduct a comparative analysis of methods for ranking vertices in graphs
- perform software implementation of methods for ranking vertices in graphs;
- to investigate the effectiveness of algorithms and methods for different input data by conducting computational experiments.
- analyze experimental data and compare with existing solutions.

**The object of research** is the process of analysis of networks for different purposes

**The subject of research** - models and methods of analysis of networks for different purposes.

**The scientific novelty** of the obtained results lies in the study of mathematical models of ranking vertices in a graph, as well as in the introduction of a system that conducts a comparative analysis of these models.

**Publications.** Vykhlyaeva A., Popenko VD Analysis of social networks // Perspectives of world science and education. Abstracts of the 9th International scientific and practical conference. CPN Publishing Group. Osaka, Japan. 2020. Pp. 375-391.

SOCIAL NETWORK, TOP, RANKING, ELECTRICAL CENTRALITY, CENTRALITY, PAGERANK.

## ЗМІСТ

### ПЕРЕЛІК УМОВНИХ

ПОЗНАЧЕНЬ.....**Ошибка! Закладка не определена.**

ВСТУП.....**Ошибка! Закладка не определена.**

1. ОГЛЯД МЕТОДІВ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ.....**Ошибка! Закладка не определена.**

1.1. Огляд наукових досліджень ..... **Ошибка! Закладка не определена.**

1.2. Методи аналізу мереж різного призначення ..... **Ошибка! Закладка не определена.**

1.3. Методи випадкових блукань та їх застосування в мережах ....**Ошибка! Закладка не определена.**

1.4. Проблеми аналізу мереж різного призначення . **Ошибка! Закладка не определена.**

Висновки за розділом ..... **Ошибка! Закладка не определена.**

2. МОДЕЛІ ТА МЕТОДИ РАНЖУВАННЯ ВЕРШИН В ГРАФАХ МЕРЕЖ  
РІЗНОГО ПРИЗНАЧЕННЯ

.....**Ошибка! Закладка не определена.**

2.1. Змістовна постановка задачі..... **Ошибка! Закладка не определена.**

2.2. Математична постановка задачі.... **Ошибка! Закладка не определена.**

2.3. Огляд методів ранжування вершин в графах .... **Ошибка! Закладка не определена.**

2.3.1. Класична центральність ..... **Ошибка! Закладка не определена.**

2.3.2. Електрична центральність ..... **Ошибка! Закладка не определена.**

2.3.3 PageRank центральність ..... **Ошибка! Закладка не определена.**

2.4. Аналіз проведених експериментів.....42

Висновки за розділом ..... 49

3. ОПИС ПРОГРАМНОГО ТА ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ..... 50

3.1. Вимоги до програмного забезпечення..... 50

3.2. Засоби розробки ..... 50

3.3. Архітектура програмного забезпечення ..... 51

3.4. Керівництво користувача..... 53

Висновки за розділом ..... 54

4. РОЗРОБКА СТАРТАП-ПРОЕКТУ.....55

4.1. Опис ідеї стартап-проекту ..... 56

4.2. Технологічний аудит ідеї проекту ..... 57



4.3. Аналіз ринкових можливостей запуску стартап-проекту .....	63
4.4. Розробка ринкової стратегії.....	64
Висновки до розділу .....	65
ВИСНОВКИ.....	66
ПЕРЕЛІК ПОСИЛАНЬ.....	67
ДОДАТОК А Графічний матеріал.....	74
Схема структурна варіантів використання.....	75
Схема структурна контекстної моделі системи.....	76
Схема структурна класів програмного забезпечення.....	77
Діаграма діяльності роботи програми.....	78
Екранні форми.....	79
Результати експериментів.....	80

## **ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ**

PR – PageRank

БД – база даних

БП – бізнес-процес

ПЗ – програмне забезпечення

ПО – предметна область

ІТ – інформаційні технології

ТЗ – технічне забезпечення

СМ – соціальна мережа

ЕЦ – електрична центральність

АСМ – аналіз соціальних мереж

## ВСТУП

В даний час аналіз соціальних мереж – один з напрямків, що найбільш інтенсивно розвивається не тільки в соціології, але і в інших гуманітарних і технічних дисциплінах. Інтерес до них пояснюється тим, що абсолютно зрозумілим стає факт особливого положення даного об'єкта дослідження, що тягне за собою новий набір пояснювальних моделей і аналітичних інструментів, які перебувають поза рамками звичайних методів дослідження - як кількісних, так і якісних. Аналіз соціальних мереж використовується для дослідження взаємодій між учасниками мережі, прогнозування їх поведінки, класифікації, моделювання інформаційних потоків в мережах. Аналіз соціальних мереж можна розуміти як дослідження людських відносин за допомогою теорії графів. Передбачення посилянь у соціальних мережах, персоналізовані методи пошуку графів, виявлення спаму у всесвітній павутині та колаборативна фільтрація в мережах рекомендацій є важливими практичними проблемами, які значною мірою покладаються на теоретичні міри подібності графів. Всі ці проблеми можуть бути охарактеризовані як ранжування сутностей у графі. Візьмемо для прикладу соціальні мережі в Інтернеті, де важливим сервісом є запропонувати друзів новачкові. Іншим прикладом є системи рекомендацій щодо кіно та музики (NetFlix, Last.fm), де користувачеві пропонуються нові фільми та музика на основі його вподобань. Пошук за ключовими словами у великих базах даних публікацій призводить до ще однієї цікавої постановки проблем, коли база даних може розглядатися як граф відношення сутності між статтями, авторами та словами. Мета - знайти документи, які "контекстуально" схожі на запит, поданий користувачем.

Одним з основних інструментів в аналізі мереж різного призначення є міра центральності, яка відповідає тому, наскільки добре розглянута вершина, розташована на шляхах, які з'єднують інші вершини. Це важливо і для додатків в транспортних задачах, а також в питаннях поширення інформації в соціальних мережах. Одним з перших визначень центральності було поняття "betweenness centrality" [6], яке було пов'язане з найкоротшими шляхами, які

проходять через розглянуту вершину. Однак, розгляд тільки найкоротших шляхів викликає серйозні заперечення, оскільки інформація може поширюватися і не обов'язково по найкоротшому шляху. Тому з'явилися міри центральності, засновані на випадкових блуканнях (PageRank) [5], і інші міри центральності, засновані на обчисленні зворотної матриці Лапласа. Останні мають хорошу аналогію з електричними мережами та законами Кірхгофа [3,4].

# **1 ОГЛЯД МЕТОДІВ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ**

## **1.1 Огляд наукових досліджень**

Сьогодні багато вчених звертаються до даного напрямку, однак серед українських дослідників мало наукових розробок у цій галузі, всі існуючі роботи за цієї темою переважно написані зарубіжними авторами, багато з них не перекладені українською або російською мовами і недоступні широкому колу читачів.

Серед найбільш актуальних проблем аналізу соціальних мереж можна назвати створення інтегральної теорії мереж, так як різноманітність підходів до їх аналізу призводить до проблеми об'єднання результатів, отриманих за допомогою кожного з підходів в рамках єдиного дослідження. Необхідно підібрати універсальний набір мір відстаней, що дозволяють визначати дистанції між елементами мережі, а також необхідно систематизувати різні заходи повноти мереж.

Стаття [1] присвячена проблемі аналізу комп'ютерних соціальних мереж. Наведено короткий опис найбільш популярних комп'ютерних соціальних мереж і перераховані окремі цікаві програмні додатки для їх аналізу. Намічено деякі можливі шляхи подальших досліджень, а саме: необхідність створення інтегральної теорії соціальних мереж, більш істотна адаптація методів обробки текстової інформації до мережевого контенту. В цій роботі [1], також, запропоновані напрями застосування випадкових блукань в соціальних мережах, Також автор зазначив, що алгоритм ранжування PageRank може застосовуватися для пошуку і класифікації сутностей і учасників в соціальній мережі. Але ця тема була не розкрита, автор не розповів, як і яким чином можна застосувати цей алгоритм для об'єднання учасників в групи відносно найбільш впливових учасників.

У статті [2] пропонується алгоритм знаходження підграфів із заданими властивостями великих графів, що описують соціальні мережі. Описується обчислювальний експеримент, що підтверджує ефективність запропонованого алгоритму. Виділення модулів (спільнот) є актуальним завданням теорії графів.

Автор статті розглядає задачу знаходження по заданому мультиграфу розбиття вершин на спільноти, при якому мінімізується задана цільова функція. Запропонована методика передбачає проведення попередньої обробки даних, яка полягає в отриманні графа на основі даного мультиграфа. Кратні зв'язки в мультиграфі замінюються на одну з вагою, рівною кількості кратних зв'язків між парою вершин. Автор дуже детально математично обґрунтовує свою точку зору. В цій роботі автор викладає про випадкові блукання в графах та показує їх роботу на прикладі соціальних мереж, автор розглядає загальну схему роботи алгоритму виділення спільнот. Єдиним недоліком роботи є те, що автор лише в одному напрямку розказує про застосування випадкових блукань в соціальних мережах.

У роботі [3] Отте проводить дослідження, що стосуються методів аналізу соціальних мереж та проблем соціального аналізу; однак немає дослідження, які стосуються сучасних та популярних проблем аналізу соціальних мереж у цій галузі. Кін та ін. у своїй праці [4] уточнили деякі проблеми АСМ, такі як аналіз настроїв та розробка думок та Інтернет виявлення спаму в своїх коротких опитуваннях, в яких вони пояснювали різні методи машинного навчання та обчислення, які використовуються для аналізу даних із багатьох соціальних платформ, таких як соціальні мережі, соціальні медіа та сайти соціальної інформації. Барб'є та Лю [5] зверталися до виявлення спільноти чи групи, розповсюдження інформації, поширення впливу, виявлення тем, індивідуальний аналіз поведінки та різні проблеми АСМ під час опису методів видобутку даних, які використовуються для отримання нової та нерозкритої інформації з отриманих даних в онлайн-середовищі соціальних медіа. Описуючи аналіз даних у соціальних мережах у 2011 році, Агарваль [6] згадував такі проблеми, як виявлення спільноти, аналіз соціального впливу та прогнозування зв'язків у соціальних мережах. Гундеча та Лю [7] у своїх дослідженнях наводили приклади реального життя, описуючи обмін даними та соціальні медіа, та визначали такі популярні проблеми аналізу соціальних мереж, як аналіз спільноти, аналіз настроїв та обмін думками, соціальні

рекомендації, моделювання впливу, інформація розповсюдження та походження, а також конфіденційність, безпека та довіра. У дослідженні про методи видобутку даних, які використовуються для АСМ, Адедоїн-Алове та ін. [8] згадали проблеми, такі як виявлення спільноти, аналіз думок, аналіз настроїв та виявлення і відстеження тем. Нанді та Дас [9] у своїй праці з важливих напрямків досліджень в соціальних мережах в Інтернеті коротко розглянули проблему аналізу соціальних мереж в Інтернеті, таких як поширення впливу, виявлення спільноти чи групи, пошук експертів, прогнозування зв'язків, системи рекомендацій, передбачення довіри та розповсюджувачів. Жюллі та ін. [10] обговорювали проблеми онлайн соціальної мережі щодо виявлення тем та виявлення впливових розповсюджувачів в онлайн соціальній мережі, ідентифікацію впливових розповсюджувачів в онлайн соціальній мережі. Гуеліл та Бухальфа [11] згадували про великі соціальні дані ЗМІ в своїх опитуваннях щодо аналізу думок та настроїв, а також обговорили проблеми аналізу соціальних мереж, такі як виявлення спільноти, виявлення спаму, прогнозування зв'язків, аналіз впливу, розповсюдження чи поширення інформації, а також конфіденційність та безпека. Пізніше вони спеціалізувалися на розробці думок та аналізі настроїв. В іншому дослідженні Агарваль [12] згадує такі проблеми, як виявлення громади, прогнозування зв'язків, поширення інформації та соціального впливу та ідентифікація експертів, а також методи видобутку даних, що застосовуються в цій галузі.

В роботі [13] автори показують, як використовувати методи випадових блукань для проектування репутаційних систем, резистентних до маніпуляцій. У статті [39] за допомогою великого числа комп'ютерних експериментів виявлено степеневий закон розподілу компонент вектора PageRank, розрахованого для web-графа, згенерованого за моделлю Баклі-Остгуса. Детально обговорено вибір алгоритмів, які використовувалися в експериментах з розрахунку вектора PageRank. Розглянуто наближену до реальності модель ранжування web-сторінок. Також обговорюються обчислювальні аспекти

запропонованої моделі в контексті раніше описаних численних способів пошуку вектора PageRank. В роботі [40] автори вивчають прогнозування гіперпосилань та проблему ефективного поєднання інформації з мережевої структури з великою кількістю вершин та ребер. У своєму дослідженні науковці розробили алгоритм, заснований на випадкових блуканнях, який природньо поєднує інформацію з мережевої структури з атрибутами вершини та ребра. Було сформульовано контрольне навчальне завдання, яке призначає значущі ребра мережи таким чином, щоб випадковий користувач швидше відвідав вершини, до яких у майбутньому будуть створені нові посилання. У роботі [41] наведено короткий огляд різних методів машинного навчання та обчислень, які використовуються в соціальних обчисленнях, спочатку вивчаючи соціальні платформи, наприклад, сайти соціальних мереж, соціальні медіа, соціальні ігри, соціальні закладки та сайти соціальних знань, де обчислювально необхідна методологія для збору, вилучення, обробки, розробки та візуалізації даних. Потім представлено опитування більш конкретних прикладів обчислювальних завдань та методик, наприклад, аналіз соціальних мереж, моделювання зв'язків та видобуток, ранжування, аналіз настроїв тощо, які використовуються на цих соціальних платформах для отримання бажаних результатів. У роботі [42] розглядається проблема оцінки заходів соціальної мережі. Зокрема, ці заходи - це середні та глобальні коефіцієнти кластеризації в мережі та кількість зареєстрованих користувачів. Належні алгоритми передбачають відсутність попередніх знань про мережу та доступ до мережі, використовуючи лише загальнодоступний інтерфейс. Більш точно, ця робота забезпечує єдиний підхід для оцінки коефіцієнтів кластеризації та новий оцінювач розміру мережі. Даний підхід полягає в тому, щоб переглянути соціальну мережу як непрямої граф і використовувати загальнодоступний інтерфейс для отримання випадкового блукання. Для оцінки коефіцієнта кластеризації по черзі тестується підключення кожного вузла у випадковому блуканні. В роботі [43] вирішується проблема пошуку спільнот у спрямованих мережах. Запропонований метод використовує вбудований PageRank, який



базується на випадкових блуканнях, щоб перетворити спрямовану мережу в непряму мережу, де інформація про ребра ефективно включена в ваги ребер. Починаючи з цієї нової непрямої зваженої мережі, раніше розроблені методи пошуку непрямого пошуку спільноти мереж можна використовувати без будь-яких змін. В статті [44] пропонується використовувати прості випадкові блукання, метод вибірки, що підтримується більшістю соціальних мереж в Інтернеті для оцінки різних властивостей великих соціальних мереж. Завдяки великомасштабному характеру соціальних мереж, оцінювачі, які отримані за схемою випадкових блукань, значно кращі, ніж рівномірні випадкові вибірки. У статті спочатку пропонується використовувати гармонічне середнє для оцінки середнього ступеня соціальної мережі. Точна оцінка середнього ступеня призводить до виявлення інших властивостей, таких як розмір популяції, неоднорідність ступенів, кількість друзів друзів. У роботі [45] проведено дослідження методів аналізу гіперпосилань, виділені основні класи ранжирування на основі зв'язків. Дано опис і характеристика алгоритму PageRank, в основу якого закладена модель випадкового блукання по веб-графу, яка використовується для обчислення ваги сторінки (коефіцієнт PageRank) як ймовірності її досяжності і алгоритму HITS-пошуку документів по заданій темі на базі гіперпосилань, в основі якого лежить ідентифікація двох наборів сторінок, які можуть бути важливими: сторінки- «авторитети» і сторінки- «концентратори». У роботі [46] представлено нове призначення методики швидшого відбору вузлів через Інтернет-соціальну мережу. Зокрема, на відміну від традиційних випадкових блукань, які чекають конвергенції розподілу вибірки до заздалегідь визначеного цільового розподілу - процесу очікування, який спричиняє високу вартість запиту, - розроблено новий алгоритм, який починається зі значно коротшого випадкового блукання, а потім попередньо оцінює показник ймовірності вибірки для вузла, взятого перед використанням вибірки прийняття-відхилення, щоб пристосувати ймовірність вибірки до заданого цільового розподілу. Алгоритм демонструє перевагу над традиційними випадковими блуканнями шляхом теоретичного аналізу та

широких експериментів над реальними світовими соціальними мережами. У роботі [47] автори доводять, що виявлення інтересів на відповідність користувача до послуг (наприклад, новин, продуктів, ігор, реклами) та прогнозування дружби для встановлення зв'язків між користувачами - це два основні завдання для соціальних мереж. У цій статті розглядається, що інформація, що міститься в мережах інтересів (тобто взаємодія користувач-сервіс) та мережах дружби (тобто підключення користувач-користувач), є дуже корельованою та взаємовигідною. Автори пропонують метод, що зв'язує користувача із зацікавленими службами та з'єднує різних користувачів із спільними інтересами, як дружба, так і будь-які ефективно пропаговані інтереси. Пропонована рамка поширення дружби та інтересів розробляє факторну модель випадового блукання на основі факторів для пояснення зв'язків дружби, і одночасно використовує сполучену модель прихованого фактора для виявлення взаємодії інтересів.

Останнім часом з'явилися роботи, пов'язані з теоретико-ігровими мірами центральності [62-63]. Зокрема, мірою центральності може служити вектор Майерсона в кооперативній грі, в якій характеристична функція являє собою число простих шляхів фіксованої довжини. У роботах [64] і [65] доведено, що вектор Майерсона можна обчислити через число простих шляхів, що проходять через дану вершину.

## **1.2 Методи аналізу мереж різного призначення.**

Існує чотири підходи щодо аналізу мереж.

**Структурний** акцентує увагу на геометричній формі та інтенсивності взаємодій (вазі ребер). Усі актори розглядаються як вершини графа, які впливають на конфігурацію ребер і інших акторів мережі. Особлива увага приділяється взаємному розташуванню вершин, центральності, транзитивності взаємодій. Для інтерпретації результатів у даному випадку використовуються структурні теорії і теорії мережевого обміну [14].

**Динамічний** – увага акцентована на змінах у мережевій структурі з часом. Вивчаються причини зникнення і появи ребер мережі; зміни структури мережі при зовнішніх діях; стаціонарні конфігурації соціальної мережі [14].

**Нормативний** – вивчає рівень довіри між учасниками, а також правила та санкції, які впливають на поведінку акторів у соціальній мережі та процеси їх взаємодій. У цьому випадку аналізуються соціальні ролі, які пов'язані з даним ребром мережі, наприклад, відношення керівника і підлеглого, дружні, родинні зв'язки. Комбінація мережевих ресурсів учасника з нормами і правилами, що діють у даній соціальній мережі, утворює його «соціальний капітал» [14].

**Ресурсний** – розглядає можливості акторів по залученню індивідуальних і мережевих ресурсів для досягнення певної мети і диференціює акторів, які знаходяться в ідентичних структурних позиціях соціальної мережі, за їх ресурсами. Як індивідуальні ресурси можуть виступати знання, престиж, багатство, етнічність, стать (гендерна ідентичність). Під мережевими ресурсами розуміються вплив, статус, інформація, капітал [14].

**Структурні методи аналізу соціальних мереж.** При структурному аналізі і аналізі поведінки зв'язків використовуються методи статистичного аналізу, визначення співтовариств, алгоритми класифікації. За допомогою структурних методів аналізу соціальних мереж досліджується взаємна поведінка вершин мережі, виходячи з припущення, що у більшості вершин мало зв'язків; досліджується, чи виникають скупчення, чи ступеня вершин розподіляються більш рівномірно; вивчається поведінка вершин при кластеризації, а також поведінка типових часових характеристик соціальних мереж, як міняється структура мережі в процесі зростання, як міняється поведінка, розподіл зв'язних компонентів графа та інше. На основі дослідження пов'язаних публікацій [1-14] було складено класифікацію та перелік структурних методів аналізу соціальних мереж, які представлені у таблиці 1.1.

Таблиця 1.1 - Структурні методи аналізу соціальних мереж

Призначення	Назви методів
-------------	---------------

методів	
Дослідження властивостей структурної позиції користувача соціальної мережі Продовження таблиці 1.1	Дослідження степеня вершини, близькості, посередництва тощо
	Методи визначення центральності вершини за посередництвом
	Методи визначення центральності вершини за близькістю
	Методи визначення центральності вершини за степенем
	Методи визначення центральності вершини за власним вектором
	Методи визначення центральності на основі законів Кірхгофа
	Теоретико-ігрові методи визначення центральності вершин в комунікаційних графах
	Методи оцінки впливовості користувача соціальної мережі через його центральність
	Методи оцінки помітності користувача соціальної мережі через його центральність та престижність
Дослідження структурних властивостей соціальної мережі	Дослідження збалансованості та транзитивності соціальних взаємодій, структурної еквівалентності акторів тощо
	Визначення централізації мережі, ступеня кластеризації, зв'язності, діаметру тощо
Визначення співтовариств (кластеризація мережі)	Ієрархічні методи: агломеративні та дивізімні
	Неієрархічні (ітеративні) методи, наприклад, алгоритм кластеризації BSP тощо
	Методи з використанням функції модульності, зокрема рандомізований алгоритм виділення

	співтовариств
Дослідження структури співтовариств та їх взаємного розташування	Аналіз зв'язних підгруп
	Ймовірнісні (стохастичні) моделі соціальних мереж
	Моделі локальної оптимізації
Дослідження та продовження таблиці 1.1 інтенсивності взаємодій	Статистичний аналіз
	Методи мережевого обміну

Одним з базових понять в аналізі мережевих структур є центральність (betweenness centrality) [66]. Перший опис центральності було здійснено в [42]. Центральність вершини - це важлива міра, що відображає те, наскільки вершина бере участь в процесі поширення інформації між іншими вершинами в графі. За допомогою центральності учасника можна визначити його помітність, а за допомогою престижності можна дізнатися і про його помітність, і про його впливовість. Дослідження центральності за ступенем допоможе визначити учасників, що контактують за великою кількістю інших учасників. Центральність та престижність визначаються одними самими методами, але для визначення центральності потрібні вхідні та вихідні зв'язки учасника, а для престижності тільки вхідні. Тому, якщо в учасника велика кількість вихідних зв'язків, це вказує на його владні функції, а якщо більша кількість вхідних зв'язків – то це вказує на його популярність. Дослідження центральності за близькістю допомагає визначити учасників, які з'єднують між собою інших учасників мережі і через яких проходить максимальна кількість коротких шляхів. За допомогою дослідження центральності за власним вектором можна також визначити учасників, які мають високі показники центральності і також контактують із великою кількістю інших учасників. Хочеться зазначити, що учасники, які мають високий показник центральності і виступають єдиним зв'язуючим компонентом між великою кількістю учасників, мають великий

вплив на цих учасників. Тому вони можуть користуватися своїм становищем для розпалювання конфліктів, дезінформації, впливу думок задля вигоди, а також і навпаки. Також цікавими є методи інтенсивності взаємодій. Саме такі методи допомагають відділити стійкі зв'язки між учасниками від випадкових та другорядних, тобто дізнатися, хто кому більше є другом, тобто дізнатися про відносини між учасниками мережі.

Як вказувалося вище, соціальний граф має структуру спільнот. Виявлення спільнот є одним з важливіших питань в аналізі соціальних мереж. Алгоритмічно цю задачу можна віднести до задачі про розділення графів. Необхідно розділити мережу на щільні регіони на основі характеру зв'язків між вершинами. Класичний підхід виявлення спільнот полягає в застосуванні до соціального графу одного з алгоритмів кластеризації. Методи кластерного аналізу діляться на наступні групи: ієрархічні і неієрархічні (ітеративні). Ієрархічні діляться на агломеративні (об'єднуючі) і дивізімні (розділяючі). Агломеративні методи застосовуються для того, щоб послідовно додавати учасників в кластер (групу), який вже існує. Коротко наведемо огляд алгоритмів, які застосовують даний метод. В *методі одиничного зв'язку* за допомогою матриці відстаней визначаємо два найближчих учасника, вони і будуть утворювати перший кластер. Далі обираємо нового учасника, який є найближчим до того учасника, який вже міститься в цьому існуючому кластері. В *методі повних зв'язків* включення нового учасника відбувається за умови, якщо відстань між учасниками не менше заданої величини. В *методі середнього зв'язку* новий учасник додається в існуючий кластер при порівнянні середнього значення міри близькості із заданою пороговою величиною. *Метод Уорда* припускає, що спочатку кожен кластер містить тільки одного учасника. Потім два найближчі кластери об'єднуються. Для них визначаються середні значення кожної ознаки і розраховується сума квадратів відхилень, надалі об'єднуються кластери, які дають найменший приріст суми квадратів відхилень [15].

Щоб об'єднати два кластери, можна застосувати такі алгоритми: *метод найближчого сусіда* (ступінь близькості оцінюється між найбільш близькими об'єктами цих кластерів), *метод дальнього сусіда* (ступінь близькості оцінюється за мірою близькості між найбільш віддаленими об'єктами кластерів), *метод середнього зв'язку* (ступінь близькості оцінюється як середня величина степенів близькості між об'єктами кластерів), *метод медіанного зв'язку* (відстань між будь-яким кластером  $S$  і новим кластером, який вийшов в результаті об'єднання кластерів  $P$  і  $Q$ , визначається як відстань від центру кластера  $S$  до середини відрізка, що сполучає центри кластерів  $P$  і  $Q$ ) [15].

За допомогою дивізімних методів в окремий кластер виділяються вершини, які мають найменші показники схожості, при чому вся мережа спочатку виступає як окремий кластер.

Окрім ієрархічних методів також варто згадати ітеративні методи. Сутність їх полягає в тому, що процес класифікації починається з завдання деяких початкових умов (кількість утворених кластерів, поріг завершення процесу класифікації і т. д.). Як і в ієрархічному кластерному аналізі, в ітеративних методах існує проблема визначення числа кластерів. Не всі ітеративні методи вимагають початкового завдання числа кластерів, але дозволяють використовувати різні алгоритми, змінюючи або число утворених кластерів, або встановлений поріг близькості для об'єднання учасників в кластери для найкращого розбиття за критеріями якості, що задаються. Ітеративні методи можуть призвести до утворення перетину кластерів, коли один учасник може одночасно належати декільком кластерам (наприклад, один користувач соціальної мережі може бути підписаним на декілька груп або каналів). Серед всіх неієрархічних методів найбільш відомий алгоритм  $k$ -середніх – швидкий кластерний аналіз.

**Нормативні методи аналізу соціальних мереж.** За допомогою нормативних методів можна досліджувати вплив та взаємовідносини між учасниками мережі. Такі методи стануть у нагоді при вирішенні наступних задач: моделювання впливу на основі інформації про учасників; моделювання

розповсюдження впливу; пошук учасників, які є найвпливовішими в процесі розповсюдження інформації та здійснення впливу на інших учасників мережі. До нормативних методів аналізу мережі ще можна віднести методи аналізу соціальних ролей та мережевого капіталу учасників мережі, бо суспільство, в якому ми живемо, дуже сильно впливає на розвиток нашої соціальної ролі та культури. При дослідженні літератури [3-9, 16], було виділено та проведено класифікацію основних нормативних методів, які наведені в таблиці 1.2.

Таблиця 1.2 - Нормативні методи аналізу соціальних мереж

Призначення методів	Назви методів
Аналіз соціальних ролей	Аналіз моделей рольової поведінки
	Аналіз дотримання соціальних норм
	Дослідження цінностей і установок, що визначають соціальну поведінку
	Аналіз мотивів соціальної поведінки
Дослідження мережевого капіталу учасників мережі	Аналіз зв'язків актора з іншими акторами, зокрема відносин довіри
	Аналіз процесів формування, акумулювання і зносу соціального капіталу акторів
	Дослідження ресурсної забезпеченості мережі
Дослідження впливу	Моделювання впливу на основі інформації про учасників
	Моделювання розповсюдження впливу
Пошук впливових учасників мережі	Пошук експертів (методи мурашиної колонії)
	Пошук брокерів (лідерів)
Аналіз рівня довіри між акторами	Дослідження рівня міжособистісної довіри
	Дослідження рівня деперсоніфікованої довіри
	Дослідження рівня генералізованої довіри
	Дослідження рівня вимушеної (обумовленої) довіри та ефективності колективних санкцій



Вважається, що в нормативному напрямку найцікавішим є дослідження вимушеної довіри. Вимушена довіра ґрунтується на здатності соціальної мережі в цілому та її впливових учасників зокрема вплинути на свого пересічного учасника, щоб змусити його дотримуватися існуючих в ній норм і правил [16].

Оскільки нормативні методи вивчають та аналізують вплив, який здійснює соціальна мережа на своїх учасників, актуальним є напрям виявлення експертів в певній області та інформаційних брокерів. Для ідентифікації експертів застосовують метод мурашиної колонії. Досить таки важливим є пошук брокерів (лідерів), адже вони виступають посередниками в соціальній мережі, які зв'язують між собою групи людей, налагоджуючи зв'язки між фахівцями і тим самим відкриваючи їм доступ до інформації.

**Динамічні методи аналізу соціальних мереж.** Динамічні методи аналізу соціальних мереж є найменш розвинутими, але все ж таки найперспективнішими. Задача динамічних методів полягає у прогнозуванні відносин між учасниками соціальної мережі та спільнот (груп), тобто прогнозування реакції учасників соціальної мережі на здійснювальний вплив. Для вирішення такого плану задач пропонується будувати різноманітні структурні і реляційні моделі. Для кращого дослідження динаміки мережі застосовують алгоритми, що поєднують у собі методи аналізу та методи візуалізації для доступного розуміння поведінки самої мережі. Також методи динамічного аналізу використовуються для прогнозування формування відносин та зв'язків між учасниками мережі. Рішенням даного питання є передбачення про з'єднання двох конкретних вершин через деякий проміжок часу. Для обчислення такої задачі використовують певні характеристики мережі, а саме кількість спільних сусідів, геодезична відстань, впливовість вершин тощо.

Для динамічного аналізу соціальних мереж використовується теорія ігор, теорія мережних ігор (розділ теорії ігор) та теорія графів. Варто зазначити, що теорія мережних ігор спеціалізується на створенні стійких зв'язків між гравцями при існуванні різних інтересів і/або різної інформованості останніх.

Ігри на соціальних мережах – ігри, в яких вершинами являються агенти – учасники соціальної мережі, а зважені дуги відображають ступень їх "довіри" один до одного або впливу один на одного. Також в даній моделі існують гравці, що здатні впливати на агентів з урахуванням їх довіри один до одного, вони здатні здійснювати цілеспрямований вплив на агентів [16]. Наведемо основні етапи дослідження ігор на соціальних мережах [17]: опис мережі та дослідження її динаміки; опис множини гравців, їх вподобань, інформованості, множин припустимих стратегій і контрольованих ними параметрів; зведення гри на мережі до тої чи іншої відомої теоретико-ігрової моделі; застосування класичного теоретико-ігрового аналізу до поставленої задачі.

В таблиці 1.3 наведено основні методи динамічного аналізу соціальних мереж, які були вивчені при аналізі супутньої літератури [1-8,16-19].

Таблиця 1.3 - Динамічні методи аналізу соціальних мереж

Призначення методів	Назва методів
Дослідження довгострокових змін між великими співтовариствами в часі	Методи з використанням теорії ігор
	Аналіз частотних моделей графів
Дослідження змін в мережі з часом	Аналіз причин появи зв'язків в мережі
	Дослідження змін структури мережі
	Візуалізація змін в структурі спільнот з часом
Прогноз формування зв'язків	Побудова структурних та реляційних моделей

**Ресурсні методи аналізу соціальних мереж.** Аналіз змісту соціальних мереж – основний напрямок дослідження ресурсних методів аналізу. Важливою задачею є визначення впливовості різних дій на учасників в соціальних мережах. Під ресурсами учасника мережі розуміються засоби, якими він регулює для впливу на інших учасників мережі. За допомогою методів аналізу соціальних мереж можна обчислити ефективність розповсюдження тієї чи іншої інформації в конкретній соціальній мережі, а також розрахувати

найбільш ефективну форму подачі цієї інформації для досягнення потрібного впливу. На основі вивчення літератури за даною тематикою [1-10] було складено таблицю 1.4 із основними ресурсними методами аналізу.

Таблиця 1.4 - Ресурсні методи аналізу соціальних мереж

<b>Призначення методів</b>	<b>Назва методів</b>
Аналіз ресурсів впливових акторів мережі	Аналіз інформаційних ресурсів
	Аналіз емоційних ресурсів
	Аналіз мережевих ресурсів
Аналіз контенту	Методи випадкових блукань
	Аналіз текстової інформації
	Аналіз мультимедіа
	Сенсорний і потоковий аналіз
Аналіз тегів	Дослідження моделей тегування
	Візуалізація тегів
	Дослідження властивостей потоків тегів
Створення рекомендацій	Колаборативна фільтрація

Аналіз текстової інформації також є дуже перспективним дослідженням в ресурсному напрямку. Як відомо, багато текстової інформації, а саме повідомлень, записів, коментарів, посилань, міститься та передається в соціальних мережах. Особливо цікавим є вивчення розстановки тегів, які описують контент мережі, а саме властивості потоків тегів, моделей тегування, семантику тегів, візуалізацію тегів. Аналіз мультимедії також можна звести до аналізу текстової інформації.

Сенсорний і потоковий аналіз. Аналізуються дані, що поступають в режимі реального часу в соціальні мережі. Багато сучасних мобільних телефонів підтримують можливість взаємодії користувачів один з одним динамічно в режимі реального часу залежно від їх місцезнаходження і статусу, наприклад, розміщення поточного місцезнаходження, транслявання потокового відео тощо.

Колаборативна фільтрація – це метод, що дозволяє користувачам позначати контент як цікавий чи нецікавий та застосувати цю інформацію для фільтрації контенту і формування рекомендацій іншим користувачам. Метод базується на пошуку учасників із схожими вподобаннями, також його можна застосувати для складання рейтингу учасників мережі або подій. Щоб порівняти людей, потрібно визначити коефіцієнт подібності на основі оцінок, які вони поставили контенту.

Виявлення потенційно шкідливого інформаційно-психологічного впливу в контенті може бути засноване на пошуку засобами машинного навчання (наприклад, нейронними мережами) певних шаблонних елементів, або пошуку методами статистичного та фрактального аналізу часто повторюваних елементів (що повторюються з метою отримання гіпнотичного ефекту) [20]. З точки зору інформаційної безпеки держави важливим є аналіз ресурсного потенціалу впливових учасників мережі з метою виявлення ресурсів, які сприяють популяризації та впливовості актора в конкретній соціальній мережі, враховуючи особливості контингенту її учасників; аналіз контенту соціальної мережі, виявлення деструктивного контенту [17].

### **1.3 Методи випадкових блукань і їх застосування в соціальних мережах.**

Розглянемо детально випадкові блукання, адже дослідження цієї роботи буде зосереджено на даних методах: інтуїція, що стоїть за їх корисністю та ефективними алгоритмами для їх обчислення, та важливі проблеми в реальному світі, де їх було застосовано. Широко популярний підхід у розробці графів та машинної літератури полягає в обчисленні близькості між вузлами за

допомогою випадкових блукань на графах: дифузія інформації від одного вузла до іншого.

Випадкові блукання забезпечують просту структуру для об'єднання інформації з ансамблів шляхів між двома вузлами. Ансамбль шляхів між двома вузлами є найважливішим компонентом багатьох популярних заходів, таких як персоналізована трансакція [21], час натискання і переміщення [22], міра Каца [23], гармонічні функції [24]. Персоніфіковані вектори рангових сторінок (PPV) використовувались для пошуку ключових слів у базах даних [25] та графу відношення-сутності [26]. Часи натискань та переходів до емпіричного періоду показали, що вони емпірично ефективні для пропозиції запитів [27], ранжування в рекомендаційних мережах [28] та проблем сегментації зображень [29]. Гармонічні функції використовувались для автоматизованої сегментації зображень та забарвлення [30, 31] та класифікації документів [32]. Обчислення цих заходів зазвичай вимагає часу  $O(n^3)$ , що є непосильним для великих графів реального світу. Алгоритми швидкого наближення для їх обчислення об'єднують ідеї наукових обчислень, баз даних та теорії графів. Ці алгоритми поширюються на розумну попередню обробку в автономному режимі [33-35], швидке обчислення в Інтернеті з розрідженою підтримкою цих мір [36-38], до гібридних алгоритмів, які використовують найкращі з обох сторін [26].

**Алгоритм PageRank.** Це – найбільш популярний алгоритм ранжування, який базується саме на методах випадкових блукань, його основний принцип полягає у пошуку та класифікації учасників та сутностей в соціальній мережі. Алгоритм PageRank моделює процес випадкового переміщення ( "серфінгу") користувача по сторінках веб-мережі, при цьому перед кожним новим кроком із заданою ймовірністю обирається одна з двох дій: користувач з рівною ймовірністю обирає одне з гіперпосилань (ребро веб-графа ), що належать поточній сторінці (вершині вебграфа), і переходить на сторінку з обраним посиланням, або користувач здійснює "телепортацію" - переміщення на іншу сторінку, обрану випадковим чином з усіх існуючих сторінок веб-мережі.

Інтуїтивно зрозуміло, що при такому переміщенні, чим більше сторінка має вхідних посилань, тим більшу кількість разів на неї буде виконаний перехід.

**Застосування методів випадкових блукань.** Методи випадкових блукань отримали високу успішність та популярність в аналізі соціальних мереж [23] та комп'ютерному зорі [48], персоналізованому пошуку графіків [49], пошуку ключових слів у графах баз даних [25, 26], виявленні спаму [50, 51]. Окремо варто зазначити, що методи випадкових блукань можуть бути також корисні для об'єднання учасників в групи навколо найбільш впливових учасників. Методи випадкових блукань показали не аби яку ефективність у передбаченні посилань. Однак основним вузьким місцем для таких методів є адаптація до мереж з різною структурою, динамікою та масштабування до розміру мережі. Випадкові блукання забезпечують популярну, інтуїтивно зрозумілу та математично принципову модель для обчислення основної міри близькості. Глибоке розуміння поведінки цих мір є ключем до ефективного використання масивного джерела мережевих даних, яке генерується з корпоративних та державних джерел щодня.

**Застосування в комп'ютерному зорі.** Поширеною технікою в комп'ютерному зорі є використання графа як зображення кадру, де два сусідніх пікселя мають міцний зв'язок, якщо вони мають подібний колір, інтенсивність або текстуру. Горелік та ін. [52] використовують середній час звернень випадкового блукання від межі об'єкта, щоб охарактеризувати форму об'єкта з силуетів. Греді та ін. [31] запровадили новий алгоритм кластеризації графів, який, як було показано, має інтерпретацію з точки зору випадкових блукань. Час звернення від усіх вузлів до призначеного вузла було порогове для отримання розділів з різними корисними теоретичними властивостями. К'ю та ін. [48] використовували кластеризацію маршрутів, які працюють на маршрутах для надійного відстеження руху багатодільних тіл та сегментації зображення. Гармонічні функції використовувались для колоризування зображень [30] та для автоматизованої сегментації зображень [31]. Колоризації передбачає додавання кольору до монохромного зображення чи фільму. Художник

коментує зображення кількома кольоровими начерками, а вказаний колір поширюється для отримання повнокольорового зображення. Це може розглядатися як проблема класифікації у багатьох класах, коли інформація про клас (кольоровий) доступна лише для кількох пікселів. Приклад сегментації використовується у визначених користувачем мітках для різних сегментів і швидко поширює інформацію для отримання високоякісної сегментації зображення. Усі перераховані вище приклади покладаються на ту саму ідею: сусідні вузли в графі повинні мати подібні мітки.

**Аналіз тексту.** Набір документів може бути представлений у графі різними способами на основі наявної інформації. Чжу та ін. [32] будують розріджений граф, використовуючи ознаку подібності між парами документів, а потім використовують підмножину міток для обчислення гармонічної функції для класифікації документів. Інший спосіб побудувати граф із документів у базі даних публікацій - це побудувати граф відношення сутності до авторів та робіт, де папери з'єднуються через цитати та співавторів. Алгоритм ObjectRank в [25] обчислює персоналізовану трансляцію сторінок для ранжирування за ключовими словами в графі, який побудовано з бази даних публікацій. Для пошуку ключових слів «серфери» починають випадкові блукання від різних об'єктів, що містять це слово. Будь-який «серфер» або переміщається випадковим чином до сусіднього вузла, або стрибає назад до вузла, що містить ключове слово. Остаточне ранжування проводиться на основі отриманого розподілу ймовірностей об'єктів в базі даних. По суті, персоналізований проміжок сторінки для кожного слова обчислюється та зберігається в автономному режимі, а під час запиту комбінується лінійно, щоб генерувати класифікацію за ключовими словами. Мей та ін. [27] використовують час звернення для пропозиції запитів. Автори будують двосторонній граф слів і URL-запитів, таким чином, що запит підключається до URL-адреси, і якщо користувач натиснув на URL-адресу, запит буде поданий до URL-адреси. Цей граф можна розглядати як ланцюг Маркова або побудувати ланцюг Маркова між вузлами запиту, використовуючи спільні URL-адреси між двома вузлами

запитів. Для даного запиту  $q$  верхні  $k$  вузли запиту в момент звернення  $q$  пропонуються як альтернативні запити. Лаферті та ін. [52] побудували двосторонній граф із слів та документів, а також використали найбільш класифіковані слова в персоналізованому перекладі сторінки із запиту для розширення запиту. Мінков та ін. [54] побудували граф зв'язків суб'єктів власників даних електронної пошти, де сутності та відносини витягуються з притаманних кому-небудь відправлених повідомлень з соціальних мереж та вмісту електронних листів. Автори використовують випадкові блукання з перезапуском для контекстного пошуку на цьому графі. Ймовірності, отримані від випадкових блукань, доповнюються глобальними характеристиками для отримання кращої прогнозованої продуктивності. У ланцюгу Маркова, де стани відповідають словам, Тутанова та ін. [55] зосередили увагу на вивченні ймовірностей переходу по ланцюгу для отримання кращого розподілу залежностей від слів. Автори визначили ряд різних типів посилань як основний перехідний розподіл. Кожен тип посилання має відповідну матрицю переходу ймовірності. Вивчення цих основних параметрів знижує обчислювальну складність, а також дозволяє авторам вивчити складні моделі випадкових блукань.

**Колаборативна фільтрація.** Бренд та ін. [28] застосовували різні заходи в результаті випадкових блукань за ненаправленими графами, щоб рекомендувати товари користувачам на основі історії їх придбань. Автори дають емпіричні докази того, що ці заходи часто малі, якщо один із вузлів має високий ступінь. Фусс та ін. [56] використовували маршрути для рекомендації фільмів у наборі даних MovieLens і показали, що цей метод працює набагато краще, ніж метод найкоротших шляхів у завданнях передбачення посилань.

Боротьба з веб-спамом. Існує ряд алгоритмів для виявлення спаму в Інтернеті, що є надзвичайно важливим завданням. Ці алгоритми можна приблизно розділити на два класи: на основі вмісту методів та на основі графів. Змістові підходи [38] зосереджуються на контенті веб-сторінки, наприклад,



кількість слів на сторінці та в заголовку сторінки, кількість тексту прив'язки, частка видимого вмісту. Все це для того, щоб відокремити спам-сторінку від неспам-сторінки. Алгоритми на основі графів розглядають структуру посилань в Інтернеті для класифікації веб-спау. TrustRank [50] використовує аналогічну ідею - персоналізоване обчислення сторінок. Розподіл перезапуску містить веб-сторінки, які, як відомо, не є спамом.

Гармонічне ранжування також було успішно використано для виявлення спау [57]. Автори будують набір вузлів-якорів і обчислюють гармонічну функцію при перезапуску. Для хороших вузлових якорів автори використовують гармонічне ранжування, а для поганих якорів вони використовують передачу вперед для ідентифікації інших вузлів з подібними мітками. Автори використовували лише гармонічне звання з однорідним набором якоря.

#### **1.4 Проблеми аналізу мереж різного призначення.**

Проблема отримання потрібних та корисних джерел була актуальною протягом усього розвитку інформаційних технологій. Дуже часто у користувачів соціальних мереж виникають проблеми із знаходженням необхідного джерела та інформації, відповідної до введеного запиту. Для вирішення таких проблем проблемами пошукові системи звичайно впроваджують такі міри: поправка введеного тексту у полі запиту, інтелектуальний пошук, випадальні допоміжні слова та популярні запити в процесі формування запиту. Але, все ж, це не забезпечує більшої ефективності в процесі пошуку. Саме вирішення проблеми видачі адекватних та корисних результатів пошуку буде розглянуто у даній роботі.

Існує ще низка проблем, які треба дослідити та вирішити при дослідженні соціальних мереж: передбачення посилань у соціальних мережах, персоналізовані методи пошуку графів, виявлення спау у всесвітній павутині, колаборативна фільтрація, точний пошук та класифікація учасників в мережі, виявлення аномалії, виявлення подій, позиції, визначення ролей, виявлення фішингу, виявлення лідерів думок, об'єднання у групи за рахунок лідерів

думок, виявлення чуток, збереження конфіденційності, виявлення причинно-наслідкових зв'язків, виявлення інтересів, виявлення шахрайства, виявлення кіберзлочинності.

### **Висновки до розділу**

Проведено аналіз існуючих методів дослідження мереж різного призначення. Наведено огляд моделей, методів та досліджень на тему аналізу мереж різного призначення. Визначені цілі та завдання магістерської дисертації.

## 2 МОДЕЛІ ТА МЕТОДИ РАНЖУВАННЯ ВЕРШИН В ГРАФАХ МЕРЕЖ РІЗНОГО ПРИЗНАЧЕННЯ

### 2.1 Змістовна постановка задачі

Аналіз соціальних мереж досліджує структуру взаємодії між соціальними сутностями. Цими соціальними сутностями можуть виступати люди, організації, міста, країни, веб-сторінки або наукові публікації. Пропонується розв'язати задачу методами ранжування вершин графу в транспортній мережі. В якості прикладу мережі транспортного характеру було обрано всім відомий київський метрополітен. Діють три лінії, експлуатаційна довжина яких становить 69,648 км, 52 станції зі трьома підземними вузлами пересадки в центрі міста.

Щоденний пасажиропотік на момент 2019 року складає 1,91 млн. Найбільш завантаженими станціями на 2011 рік є «Вокзальна» (68,3 тис. осіб/добу), «Лісова» (66,9 тис. осіб/добу), «Почайна» (53,3 тис. осіб/добу), «Лівобережна» (52,6 тис. осіб/добу), «Дарниця» (51,8 тис. осіб/добу) та «Мінська» (50,8 тис. осіб/добу). Найменш завантаженими станціями є «Дніпро» (3,0 тис. осіб/добу), «Червоний хутір» (5,0 тис. осіб/добу), «Славутич» (6,3 тис. осіб/добу), «Гідропарк» (6,9 тис. осіб/добу) та «Вирлиця» (7,4 тис. осіб/добу).

У 2018 році найбільш завантаженою була станція «Лісова». За 2019 рік найбільш завантаженою стала станція «Академмістечко», якою скористалося більше 21,3 млн пасажирів, найменший пасажиропотік зафіксовано на станції «Дніпро», яка прийняла 938,6 тисяч пасажирів. Загалом Святошинсько-Броварська лінія з 18 станцій перевезла 203 млн 176 тис. пасажирів. Другою за завантаженістю стала Оболонсько-Теремківська лінія, якою скористалося 173,2 млн пасажирів. Максимальний пасажиропотік на цій лінії зафіксовано на станції «Мінська» (понад 17 млн осіб/рік), мінімальний — «Поштова площа» (4,9 млн пасажирів). Сирецько-Печерською лінією за 2019 рік скористалися майже 118,8 млн пасажирів. Найбільші пасажиропотоки на зеленій лінії зафіксовано на станціях «Лук'янівська» (14 млн осіб/рік)

та «Позняки» (14,8 млн осіб/рік), найменше на цій лінії користувалися станцією «Червоний хутір» (1,8 млн осіб/рік). Всього за 2019 рік «Київський метрополітен» перевіз 495 млн 300 тис. пасажирів, що на 700 тисяч менше, ніж 2018 року.

На рисунку 2.1 зображено гілки київського метрополітену та рух швидкісного трамваю.



Рисунок 2.1 – Схема маршрутів київського метро та схема руху швидкісних трамваїв.

Побудуємо граф на основі київського метрополітену та визначимо за допомогою методів обчислення класичної центральності, електричної центральності та PageRank найбільш навантажені станції метро та порівняємо роботу вищезгаданих алгоритмів.

## 2.2 Математична постановка задачі

Будь-яку мережу можна представити у вигляді графу  $G(V, E)$ , де  $V = \{v_0, v_1 \dots v_N\}$  – множина вершин,  $v_0, v_1 \dots v_N$  – учасники мережі;  $E$  – множина ребер  $\{(v_i, v_j) | i \neq j\}$ , що відповідають наявності відносин.  $ij$ -й елемент матриці суміжності  $A$  невід'ємний і позначає вагу на перетині  $i, j$  та дорівнює нулю, якщо ребра не існує. У невагомих графах вага будь-якого ребра дорівнює 1, тоді як у зваженому графі він може бути будь-яким додатним числом.  $D$  – діагональна матриця  $n \times n$ , де  $D_{ii} = \sum_j A_{ij}$ . Позначимо  $D_{ii}$  ступенем  $d(i)$  вузла  $i$ , лапласійська  $L$  від  $G$  визначається як  $D - A$ . Це широко використовується в алгоритмах машинного навчання як регулятор ([24, 58]).

$P = p_{ij}, i, j \in V$  позначає матрицю ймовірностей переходу так, що  $P_{ij} = A_{ij}/D_{ij}$ , якщо  $(i, j) \in E$  і 0. Випадкове блукання на цьому графі – ланцюг Маркова [22] з ймовірністю переходу, визначеною цією матрицею. Для неорієнтованого графа  $A$  є симетричною і  $L$  – симетрична позитивна напіввизначеність. Позначимо набір сусідів вузла  $i$  через  $N(i)$ . Для невзваженого графа  $d(i)$  – розмір цієї околиці (спільноти). Для спрямованого графа набір вузлів, що мають ребро до вузла  $i$ , позначається як  $I(i)$ , а напівступінь входу позначається  $d^-(i)$ . Аналогічно множина зовнішніх сусідів позначається  $O(i)$ , а напівступінь виходу позначається  $d^+(i)$ . Для зважених графів зваженими є як напівступінь входу, так і виходу.

У випадкових блуканнях, якщо вузол  $v_0$  обраний із розподілу  $x_0$ , то розподіли  $x_0, x_1, \dots$  взагалі відрізняються один від одного. Однак якщо  $x_t = x_{t+1}$ , то говоримо, що  $x_t$  – стаціонарний розподіл для графа. Відповідно до теореми Перрона-Фробеніуса, існує унікальний стаціонарний розподіл, якщо  $P$  незвідна і аперіодична.

## 2.3 Огляд методів ранжування вершин в графах

### 2.3.1 Класична центральність

Нехай дано граф  $G = (V, E)$ , де  $V$  – множина вершин,  $E$  – множина ребер. Позначимо число вершин як  $n = |V|$ , а число ребер як  $m = |E|$ . Тоді центральність вершини  $v \in V$  визначається наступною формулою:

$$c_B(v) = \frac{1}{n_B} \sum_{s,t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}, \quad (2.1)$$

де  $\sigma_{s,t}$  - число геодезичних шляхів між вершинами  $s \in V$  і  $t \in V$ ,  $\sigma_{s,t}(v)$  – число геодезичних шляхів між вершинами  $s$  та  $t$ , які проходять через вершину  $v$ . Коефіцієнт нормування  $n_B$  дорівнює  $n_B = (n-1)(n-2)$ , якщо вершина  $v$  не може бути початковою  $s$  або кінцевою  $t$  вершинами, і  $n_B = n(n-1)$  навпаки. Найкраща обчислювальна складність алгоритму пошуку центральності дорівнює  $O(mn)$ .

### 2.3.2 Електрична центральність

Для визначення центральності можна використовувати методи кооперативної теорії ігор. Тут вершини графа представляють гравців, а ребра представляють зв'язок між гравцями. Гравці можуть кооперуватися тільки якщо вони пов'язані, тобто кооперація обмежена неорієнтованим графом. Далі задається характеристична функція, що визначає виграш коаліції. Таким чином, задається комунікаційна гра, в якій, визначивши вектор Майерсона [67], можна визначити центральність всіх вершин.

Відомо, що пошук вектора Майерсона має високу обчислювальну складність. В [68] була запропонована спеціальна характеристична функція, яка істотно спростила процес обчислення вектора Майерсона.

В [69-70] міра центральності обчислюється на основі моделі електричного ланцюга. У даній моделі граф розглядається як електричний ланцюг з ідеальними елементами, де кожне ребро має якусь пропускну здатність (значення, зворотне опору), а вершини графа є її вузлами. Для пошуку міри центральності в моделі електричного ланцюга використовуються закони Кірхгофа.

Для цього електричний ланцюг заземлюється в деякій вершині  $t$  і подається електричний струм в деякій вершині  $s$ . В [69] струм подається в деякій єдиній вершині  $s$ , а також в єдиній вершині  $t$  мережу заземлюється. Мірою центральності вершини  $v$  виступає середня величина струму, що проходить через вершину  $v$  по всіх можливих параметрах  $s$  і  $t$ . Таким чином, в

моделі електричного ланцюга при розрахунку міри центральності враховуються не тільки геодезичні шляхи.

Найкраща обчислювальна складність відомого на поточний момент алгоритму пошуку міри центральності в моделі електричного ланцюга дорівнює  $O(I(n-1) + mn \log n)$ , де  $I(n-1)$  – складність обчислення зворотної матриці розмірності  $n-1$ .

В [71] розглядається модель електричного ланцюга, де пропускна здатність всіх ребер графа приймається рівною деякому постійному параметру  $\alpha$ . На відміну від [69] в моделі [71] в електричного ланцюга штучно вводиться  $n+1$  заземлена вершина, і кожна вершина мережі з'єднана з нею ребром з пропускною спроможністю  $1-\alpha$ . Ці зміни дозволили істотно зменшити обчислювальну складність алгоритму. Крім того, це дало можливість застосувати стохастичні методи в обчисленні міри центральності в графах з великою розмірністю. Але дана модель була передбачена тільки для незважених графів.

Нехай задано зважений граф  $G = (V, E, W)$ , де  $V$  – множина вершин,  $E$  – множина ребер,  $W$  –

$$W(G) = \begin{pmatrix} 0 & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & 0 & \dots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \dots & 0 \end{pmatrix},$$

матриця ваг:

де  $w_{i,j} \geq 0$  – вага ребра між вершинами  $v_i$  і  $v_j$ ,  $n = |V|$  – кількість вершин. Якщо вершини  $v_i$  і  $v_j$  несуміжні, то  $w_{i,j} = 0$ . Якщо  $G$  – неорієнтований граф, то  $w_{i,j} = w_{j,i}$ .

Позначимо  $D(G)$  діагональну матрицю виду:

$$D(G) = \begin{pmatrix} d_{v_1} & 0 & \dots & 0 \\ 0 & d_{v_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{v_n} \end{pmatrix},$$

де  $d_{v_i} = \sum_{j=1}^n w_{i,j}$  - сума ваг ребер інцидентних вершині  $v_i$  в графі  $G$ .

Визначення. Матриця  $L(G)$  називається матрицею Кирхгофа (Laplacian matrix) зваженого графа  $G$ , якщо

$$L(G) = D(G) - W(G) = \begin{pmatrix} d_{v_1} & -w_{1,2} & \dots & -w_{1,n} \\ -w_{2,1} & d_{v_2} & \dots & -w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -w_{n,1} & -w_{n,2} & \dots & d_{v_n} \end{pmatrix}. \quad (2.2)$$

$$\begin{pmatrix} d_{v_1} + \delta & -w_{1,2} & \dots & -w_{1,n} & -\delta \\ -w_{2,1} & d_{v_2} + \delta & \dots & -w_{2,n} & -\delta \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -w_{n,1} & -w_{n,2} & \dots & d_{v_n} + \delta & -\delta \\ -\delta & -\delta & \dots & -\delta & \delta n \end{pmatrix}.$$

Нехай граф  $G'$  отриманий з графа  $G$  шляхом додавання додаткової  $v_{n+1}$  вершини, пов'язаної з усіма вершинами графа  $G$  ребрами з постійною (2.3) провідністю  $\delta$ . Таким чином, отримаємо наступну матрицю Кирхгофа графа  $G'$ :

$$L(G') = D(G') - W(G') =$$

Припустимо, що одиниця електричного струму подається в деяку вершину  $s \in V$ , а вершина  $v_{n+1}$  заземлена. Нехай  $\varphi_v^s$  позначає абсолютний потенціал в вершині  $v$  за умови, що джерело електричного струму підключено до вершини  $s$ . Тоді вектор абсолютних потенціалів  $\varphi^s(G') = [\varphi_{v_1}^s, \dots, \varphi_{v_n}^s, \varphi_{v_{n+1}}^s]^T$  в вершині графу  $G'$  визначається наступною системою рівнянь (правила Кіргофа) [72]:

$$\varphi^s(G') = L(G')^{-1} b_s',$$

де  $b_s'$  - вектор-стовпець, який складається з  $n + 1$  елементів, значення яких дорівнює:

$$b_s'(v) = \begin{cases} 1 & v = s \\ 0 & \text{інакше} \end{cases} \quad (2.5)$$

Так як матриця Кіргофа є виродженою (2), прийmemo в вершині  $v_{n+1}$  абсолютний потенціал рівним 0.



Тоді з (3) отримаємо:

$$\tilde{\varphi}^s(G') = \tilde{L}(G')^{-1}b'_s, \quad (2.6)$$

де  $\tilde{\varphi}^s(G')$  і  $b'_s$  отримані з  $\varphi^s(G')$  і  $b'_s$  шляхом видалення елементів, які відповідають вершині  $v_{n+1}$ , а  $\tilde{L}(G')$  отриманий з  $L(G')$  шляхом видалення рядків і стовпців, які відповідають вершині  $v_{n+1}$ . Відзначимо, що при цьому з  $\varphi^s(G')$  і  $b'_s$  видаляються елементи, які дорівнюють 0.

Нескладно побачити, що

$$\tilde{\varphi}^s(G') = [D(G) - W(G) + \delta I]^{-1}b_s, \quad (2.7)$$

де  $I$  – одинична матриця розмірності  $n$ .

Так як абсолютні потенціали можуть бути визначені з точністю до постійного доданку, то абсолютні потенціали  $\tilde{\varphi}^s(G')$  можна прийняти в якості абсолютних потенціалів в вершинах графа  $G$ , тобто

$$\tilde{\varphi}^s(G) = [L(G) + \delta I]^{-1}b_s.$$

Подано (7) в наступному вигляді:

$$\begin{aligned} \tilde{\varphi}^s(G) &= [(D(G) + \delta I) - W(G)]^{-1}b_s \\ &= [I - (D(G) + \delta I)^{-1}D(G)D^{-1}(G)W(G)]^{-1} \times (D(G) + \delta I)^{-1}b_s. \end{aligned}$$

Матриці  $(D(G) + \delta I)^{-1}$  і  $(D(G) + \delta I)^{-1}D(G)$  діагональні з елементами на діагоналі  $\frac{1}{d_i + \delta}$  і  $\frac{d_i}{d_i + \delta}$ ,  $i = 1, \dots, n$ , визначимо їх як  $D_1$  і  $D_2$ , відповідно. Матриця  $D^{-1}(G)W(G)$  стохастична, позначимо її  $P$ . Тоді

$$\tilde{\varphi}^s(G) = [I - D_2P]^{-1}D_1b_s = \sum_{k=0}^{\infty} (D_2P)^k D_1b_s. \quad (2.8)$$

З (8) випливає, що потенціал можна обчислити рекуррентно у вигляді:

$$\tilde{\varphi}_{k+1}^s(G) = D_2P\tilde{\varphi}_k^s(G) + D_1b_s, \quad \tilde{\varphi}_0^s(G) = 0.$$

Струм, що протікає через ребро  $e = (v_i, v_j)$ , відповідно до закону Ома дорівнює  $x_e^s = \left| \varphi_{v_i}^s - \varphi_{v_j}^s \right| \cdot w_{i,j}$ .

Величину струму, що протікає через вершину  $v$ , можна знайти, визначивши суму струмів на ребрах, інцидентних їй, і поділити отриману суму на 2. Відзначимо, що необхідність в розподілі на 2 виникла в зв'язку з тим, що

при підсумовуванні всіх струмів на ребрах, інцидентних вершині, враховується електричний струм як входить до вершини, так і виходить з неї.

Відповідно величину струму, що протікає через деяку вершину  $v$ , за умови, що джерело струму знаходиться в вершині  $s$ , можна знайти за формулою:

$$x^s(v) = \frac{1}{2}(b_s(v) + \sum_{e:v \in e} x_e^s),$$

$$\text{де } b_s(v) = \begin{cases} 1 & v = s \\ 0 & \text{інакше} \end{cases}.$$

Відповідно величину струму, що протікає через деяку вершину  $v$ , за умови, що джерело струму знаходиться в вершині  $s$ , можна знайти за формулою:

$$CF_\delta(v) = \frac{1}{n} \sum_{s \in V} x^s(v). \quad (2.10)$$

Обчислювальна складність алгоритму пошуку міри центральності для зважених графів (10) дорівнює складності пошуку зворотної матриці розмірності  $n$ , тобто  $O(n^3)$ .

### *Деталізований алгоритм пошуку електричної центральності*

**Input:** electrical network  $N = (G; c)$  with vertices  $v_1, \dots, v_n$   
**Output:** current-flow betweenness  $c_{CB} : V \rightarrow \mathbb{R}_{\geq 0}$

```

begin
   $c_{CB} \leftarrow 0$ 
   $C \leftarrow \begin{pmatrix} 0 & 0^T \\ 0 & L^{-1} \end{pmatrix}$ 
  for  $e \in E$  do
     $F_{e\bullet} \leftarrow (BC)_{e\bullet}$ 
    sort row  $F_{e\bullet}$  in non-increasing order
     $\text{pos}(e, v) \leftarrow \text{rank of } F_{ev} \text{ in sorted row } F_{e\bullet}$ 
    for  $i = 1, \dots, n$  do
      increase  $c_{CB}(\text{source}(\vec{e}))$  by  $(i - \text{pos}(e, v_i)) \cdot F_{ev_i}$ 
      increase  $c_{CB}(\text{target}(\vec{e}))$  by  $(n + 1 - i - \text{pos}(e, v_i)) \cdot F_{ev_i}$ 
    for  $i = 1, \dots, n$  do
       $c_{CB}(v_i) \leftarrow (c_{CB}(v_i) - i + 1) \cdot 2/n_B$ 
end

```

### 2.3.3 PageRank-центральність

Pagerank [59] - широко відомий алгоритм для ранжування веб-сторінок. Якщо всесвітня павутина розглядається як граф, то pagerank сторінки визначається як розподіл, який задовольняє такому лінійному рівнянню (2.11):

$$\mathbf{v} = (1 - \alpha)P^T \mathbf{v} + \frac{\alpha}{n} \mathbf{1}, \quad (2.11)$$

де  $\alpha$  - ймовірність перезапустити випадкове блукання на заданому кроці. Зауважемо, що коефіцієнт перезавантаження  $\alpha$  відіграє дуже важливу роль. Теорема Перрона-Фробеніуса вказує, що стохастична матриця  $P$  має унікальний головний власний вектор, якщо вона незвідна і аперіодична. Випадковий перезапуск гарантує, що, за цією моделлю: 1) всі вузли доступні від усіх інших вузлів, і 2) ланцюг Маркова є аперіодичним. Також, ймовірність перезапуску робить другу за величиною власну величину верхньою межею  $\alpha$  [60]. Простіше кажучи, сторінки, на яких, швидше за все, з'явиться випадковий відвідувач, мають великі значення в стаціонарному розподілі ланцюга Маркова, який з ймовірністю  $\alpha$  випадковим чином переходить відповідно до структури зв'язку в Інтернеті та з вірогідністю  $1-\alpha$  телепортується відповідно до вектора телепортаційного розподілу  $\mathbf{v}$ , де  $\mathbf{v}$  зазвичай рівномірний розподіл на всі сторінки. Узагальнюючи, замінемо поняття "перехід відповідно до структури зв'язку мережі" на "перехід відповідно до стохастичної матриці  $P$ ."

Проте, через те, що граф гіперпосилань реальної мережі Веб не завжди зв'язний та має тупикові вершини, у формулу випадкового блукання доводиться запровадити можливість випадкової телепортації на іншу вершину. Таким чином, формула обчислення розподілів знаходження  $\mathbf{v}'$  матиме вигляд:

$$\mathbf{v}' = \alpha M \mathbf{v} + (1 - \alpha) \frac{\mathbf{e}}{n}, \quad (2.12)$$

- $\alpha$  — константа, що зазвичай має значення в проміжку 0.8...0.9;
- $\mathbf{e}$  — одиничний вектор (розміром  $n$ , всі елементи якого дорівнюють 1);
- $n$  — кількість індексованих сторінок, відповідно — кількість вершин графу;
- $\alpha M \mathbf{v}$  — моделює випадок, коли з ймовірністю  $\alpha$  мандрівник вирішує обрати вихідне посилання з поточної сторінки;
- $(1 - \alpha) \frac{\mathbf{e}}{n}$  — вектор, кожен елемент якого дорівнює  $(1 - \alpha)/n$  і який моделює появу користувача на будь-якій сторінці з ймовірністю  $(1 - \alpha)$ .

PageRank надає демократичний погляд на важливість веб-сторінок. Однак за допомогою нерівномірної щільності ймовірності перезапуску можемо

створити персоналізований вигляд відносної важливості вузлів. Основна ідея - почати випадкове блукання від вузла  $i$ ; на будь-якому кроці хода переміщається випадковим чином до сусіда поточного вузла з ймовірністю  $1 - \alpha$  і скидається на початковий вузол  $i$  з ймовірністю  $\alpha$ . Стаціонарний розподіл цього процесу - це персоналізований PageRank вузол  $i$ . В результаті стаціонарний розподіл локалізується навколо стартового вузла. Це також називається «укорінений PageRank» [61].

Якщо узагальнимо стартовий вузол  $i$  до початкового розподілу  $r$ , вектор персоналізації буде заданий:

$$\mathbf{v} = (1 - \alpha)P^T \mathbf{v} + \alpha r. \quad (2.13)$$

Для отримання персоналізації вузла  $i$ , розподіл перезавантаження встановлений на вектор, де елемент  $i^{th}$  встановлений на 1, а всі інші записи встановлені на 0.  $P^T v$  - це розподіл після одного кроку випадкового блукання з  $v$ . Вищезазначене визначення означає, що персоналізований PageRank може бути обчислений шляхом розв'язання великої лінійної системи, що включає матрицю переходу  $P$ .

Алгоритму PageRank обчислюється за такою формулою:

$$PR(A) = (1 - d) + d \left( \frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right), \quad (2.14)$$

де  $PR(A)$ - вага сторінки  $A$  (те, що ми шукаємо),  $d$  - коефіцієнт затухання, який зазвичай встановлюють рівним 0,85,  $PR(T_1)$  – вага PageRank сторінки, що посилається на сторінку  $A$ ,  $C(T_1)$  – кількість посилань із цієї сторінки,  $\frac{PR(T_n)}{C(T_n)}$  – для кожної сторінки, яка вказує на сторінку  $A$ .

## 2.4 Аналіз проведених експериментів

Було проведено серію експериментів для обчислення класичної центральності, електричної центральності та PageRank-центральності для усіх вершин (станцій) графу (київського метрополітену). Результати роботи алгоритмів було порівняно з реальними статистичними даними, а саме з пасажиропотоком в метро. Дані взяті з офіційного сайту метрополітену (<http://www.metro.kiev.ua/>). Згідно 2019 року кожного дня метро перевозило у

середньому 1 356 986 людей. Також відомі дані за кожен день по кожній станції: Героїв Дніпра (29,3 тис.), Золоті ворота (20,6 тис.), Палац спорту (24,7 тис.), Площа Льва Толстого (25,4 тис), Майдан Незалежності (32,0 тис), Хрещатик (34,0 тис), Театральна (15,6 тис.), Політехнічний інститут (30,6 тис), Арсенальна (21,2 тис), Університет (20,2 тис.), Вокзальна (54,3 тис), Святошин (25,4 тис), Лівобережна (46,0 тис), Дарниця (43,7 тис), Лісова (55,1 тис), Бориспільська (16,7 тис), Червоний хутір (4,9 тис), Вирлиця (6,9 тис), Позняки (40,7 тис), Академмістечко (58,4 тис), Мінська (45,5 тис), Оболонь (33,0 тис), Почайна (41,2 тис), Либідська (28,6 тис), Олімпійська (31,7 тис), Контрактова площа (42,2 тис), Поштова площа (13,5 тис), Тараса Шевченка (15,5 тис), Житомирська (37,8 тис), Чернігівська (38,0 тис), Гідропарк (4,8 тис), Нивки (24,4 тис), Берестейська (22,4 тис), Шулявська (22,2 тис), Дніпро (2,5 тис), Харківська (25,6 тис), Осокорки (21,6 тис), Славутич (7,4 тис), Видубичі (31,1 тис), Дружби народів (22,8 тис), Печерська (24,6 тис), Кловська (12,1 тис), Лук'янівська (38,6 тис), Дорогожичі (15,9 тис), Сирець (11,6 тис).

Найбільш завантаженою, за підрахунками кількості поїздок, виявилася станція «Лісова». Також серед найпопулярніших – «Академмістечко» та «Вокзальна». Натомість станцією «Дніпро» користуються в 20 разів рідше. Загалом найбільший пасажиропотік – на червоній гілці метрополітену.

На синій лінії максимальний пасажиропотік зафіксували на станції «Мінська», а найменший – на «Поштовій площі». Найменш завантажена – зелена гілка. Тут найпопулярніші станції – «Позняки» і «Лук'янівська». Найменше користуються станцією «Червоний хутір».

На рисунку 2.2 наведено графік результатів обчислення класичної центральності для всіх вершин графу київського метрополітену. На графіку видно, що найбільші показники центральності мають наступні станції: Академмістечко, Лісова, Мінська, Іподром, Майдан Незалежності та Хрещатик.

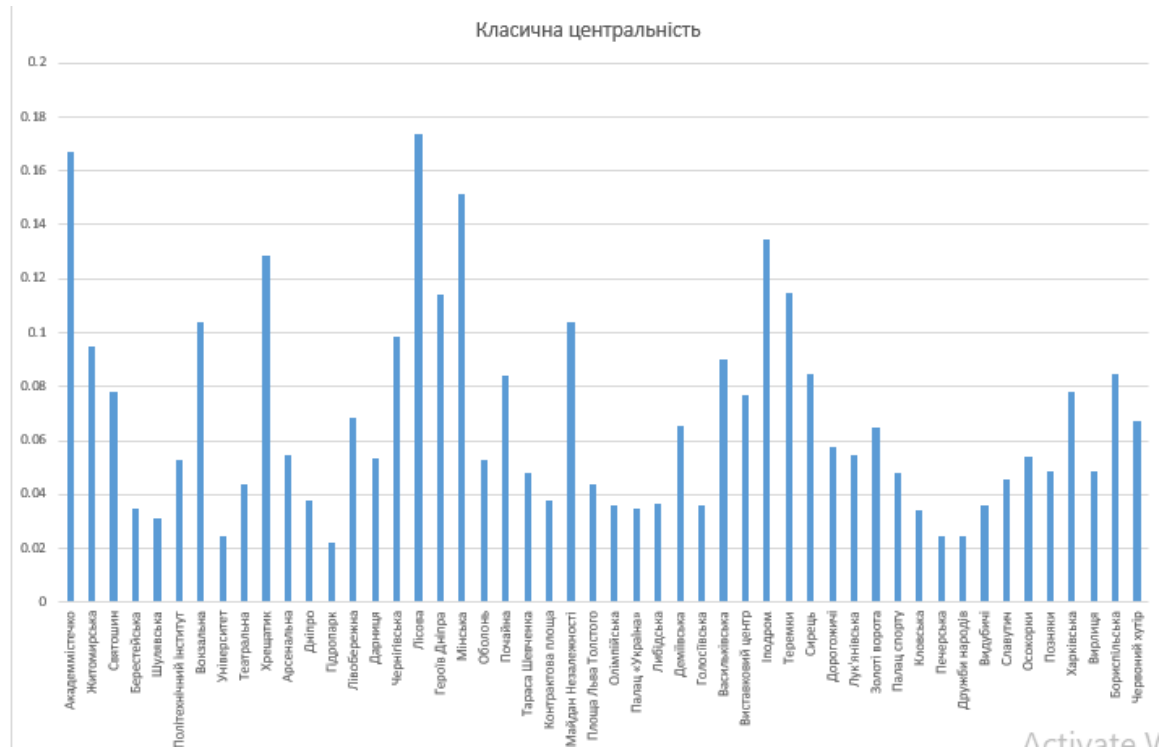


Рисунок 2.2 – Графік результатів обчислення класичної центральності для всіх вершин графу київського метрополітену.

На графіку видно, що найбільші показники PageRank-центральності мають наступні станції: Лісова, Мінська, Арсенальна, Площа Льва Толстого, Майдан Незалежності, Чернігівська, Політехнічний Інститут.

На рисунку 2.3 наведено графік результатів обчислення PageRank-центральності для всіх вершин графу київського метрополітену.

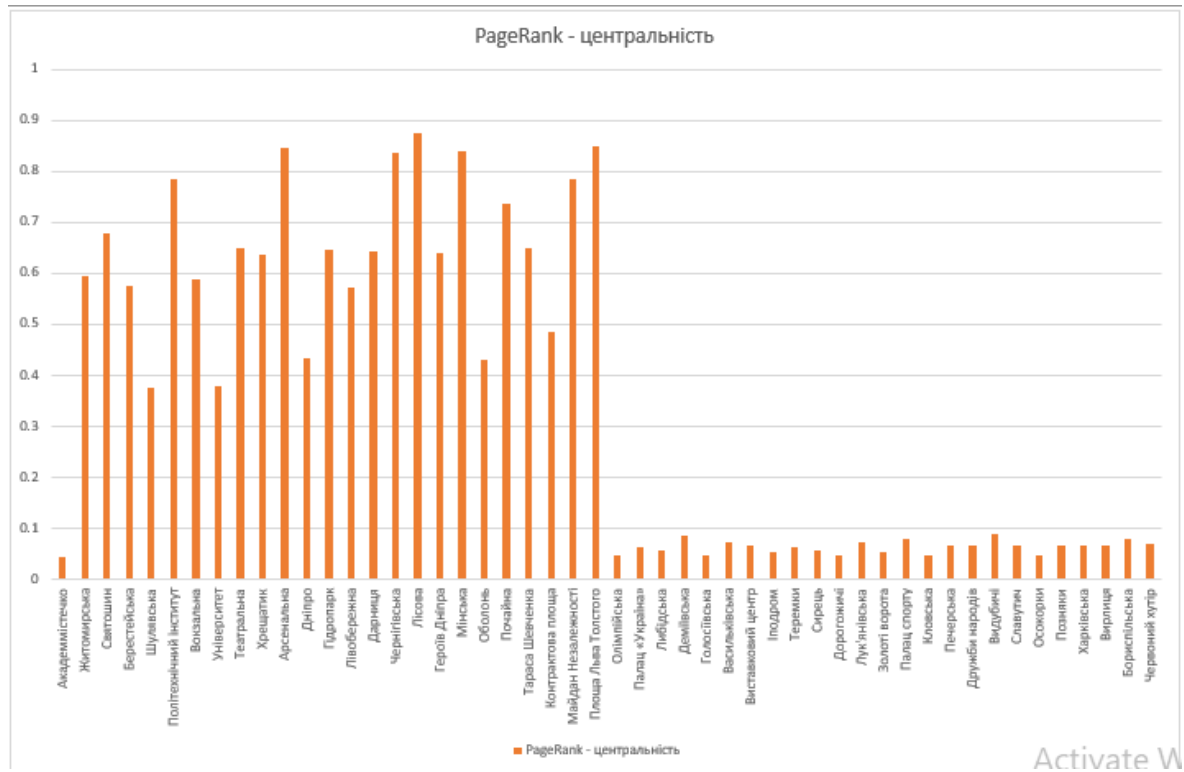


Рисунок 2.3 – Графік результатів обчислення PageRank-центральності для всіх вершин графу київського метрополітену.

На рисунку 2.4 наведено наведено графік результатів обчислення електричної центральності для всіх вершин графу київського метрополітену. Лідерські позиції мають: Академістечко, Лісова, Мінська, Вокзальна, Хрещатик та Майдан Незалежності.

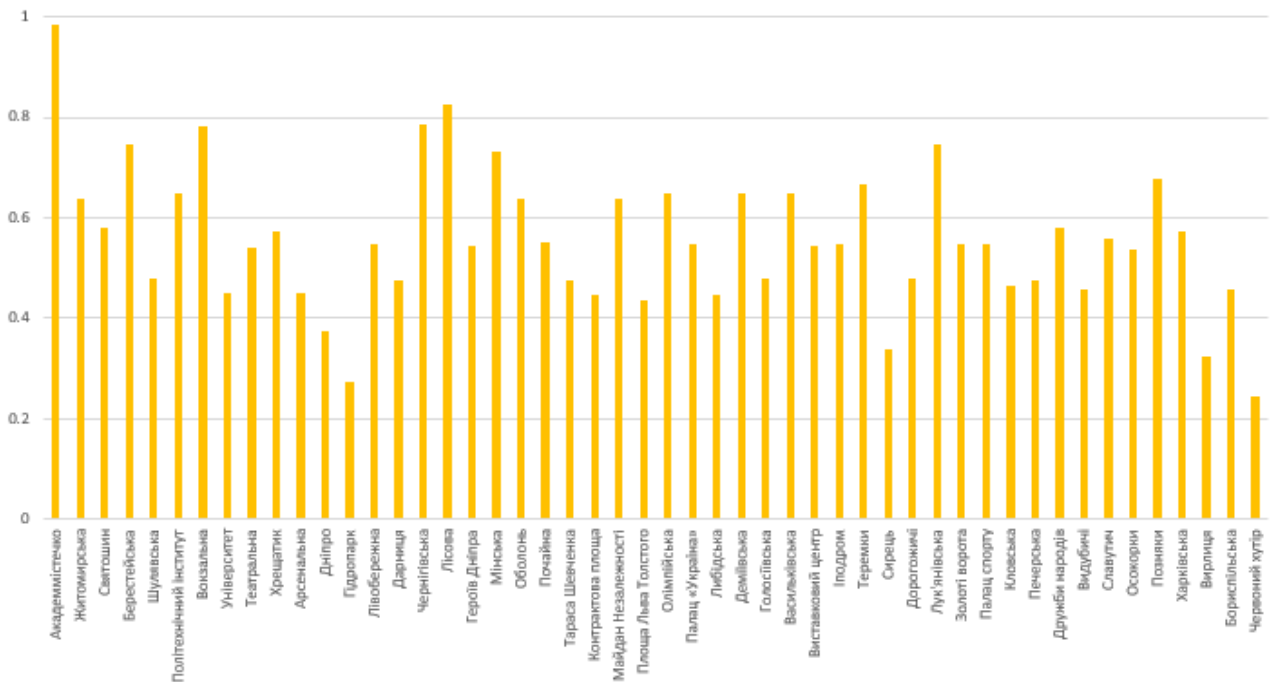


Рисунок 2.4 – Графік результатів обчислення електричної центральності для всіх вершин графу київського метрополітену.

В таблиці 2.1 представлені результати обчислення класичної, електричної та PageRank-центральності.

Таблиця 2.1 – Результати обчислення центральностей метрополітену

Назва станції	Класична центральність $\delta = 1$	PageRank – центральність $\alpha = 0.85$	Електрична центральність $\delta = 1$
Академмістечко	0.16695	0.04478	0.98543
Житомирська	0.09478	0.592942	0.638482
Святошин	0.07829	0.678372	0.578438
Берестейська	0.03472	0.573829	0.747389
Шулявська	0.03121	0.375897	0.477825
Політехнічний інститут	0.05271	0.783983	0.648390
Вокзальна	0.10389	0.587874	0.783489
Університет	0.02482	0.378927	0.448927
Театральна	0.04374	0.649374	0.539849
Хрещатик	0.12844	0.634899	0.573857
Арсенальна	0.05467	0.846367	0.448930
Дніпро	0.03748	0.433478	0.373828
Гідропарк	0.02213	0.646434	0.274839
Лівобережна	0.06842	0.573378	0.547378
Дарниця	0.05322	0.643820	0.475284
Чернігівська	0.09873	0.837478	0.783848
Діброва	0.17357	0.874389	0.824583
Героїв Дніпра	0.11398	0.638304	0.542949
Мінська	0.15130	0.839020	0.730208
Оболонь	0.05279	0.430348	0.639430
Почайна	0.08378	0.737439	0.549583
Тараса Шевченка	0.04768	0.650393	0.474727



Контрактова площа	0.03747	0.483920	0.447897
Майдан Незалежності	0.10378	0.784289	0.637482
Площа Льва Толстого	0.04374	0.847349	0.434789
Олімпійська	0.03575	0.0463088	0.647372
Палац «Україна»	0.03464	0.0638478	0.546286
Либідська	0.03678	0.0573899	0.447899
Деміївська	0.06543	0.0843848	0.647892
Голосіївська	0.03568	0.0476349	0.477294
Васильківська	0.08993	0.0738434	0.648597
Виставковий центр	0.07678	0.0648382	0.543856
Іподром	0.13478	0.0538329	0.547378
Теремки	0.11490	0.0638483	0.667388
Сирець	0.08467	0.0573468	0.338481
Дорогожичі	0.05788	0.0473846	0.477930
Лук'янівська	0.05478	0.0738480	0.747331
Золоті ворота	0.06478	0.0547834	0.547399
Палац спорту	0.04788	0.0778376	0.547372
Кловська	0.03432	0.0467838	0.464838
Печерська	0.02482	0.0658938	0.475777
Дружби народів	0.02446	0.0647891	0.578368
Видубичі	0.03582	0.0894799	0.457583
Славутич	0.04567	0.0674892	0.557588
Осокорки	0.05371	0.0455993	0.538532
Позняки	0.04829	0.0663728	0.678738

Продовження таблиці 2.1

Харківська	0.07820	0.0666372	0.573857
Вирлиця	0.04858	0.0647383	0.323589
Бориспільська	0.08472	0.0792848	0.457892
Червоний хутір	0.06735	0.0684029	0.245738

Результати ранжування електричною центральністю підтверджуються статистичними даними з офіційного сайту київського метрополітену за 2019 рік. Отже, за результатами обчислення електричної центральності найбільш високі показники отримали станції Академістечко, Лісова, Мінська, а низькі показники – Вирлиця, Червоний Хутір, Сирець та Гідропарк.

Результати ранжування PageRank дають відмінні результати від статистичних даних – станції метро Площа Льва Толстого, Арсенальна, Політехнічний Інститут мають вищі показники, аніж дійсно найбільш навантажені станції згідно пасажиропотіку за день. І навпаки, такі станції як Позняки, Теремки, Лук'янівська, які за статистичними даними мають пасажиропотік вище за середнє – отримали доволі низькі показники.

За результатами ранжування класичною центральністю найбільш високі показники отримали Академістечко, Лісова, Мінська, які підтверджуються статистичними даними з офіційного сайту київського метрополітену. Але також станції, які за статистичними даними мають найнижчі показники за пасажиропотіком не співпали за результатами обчислення ранжування класичною центральністю.

Отже, можна зробити висновок, що метод ранжування за електричною центральністю є найбільш ефективним для визначення центральності вершини в графі.

### **Висновки до розділу**

У розділі наведено змістовну та математичну постановку задачі пошуку центральності вершини в графі. Проведено огляд існуючих методів розв'язання задачі.

Було запропоновано та розроблено 3 алгоритми пошуку центральності вершини в графі, а саме класична центральність, електрична центральність та

PageRank. Описано схеми алгоритмів та проведено обчислювальні експерименти. Виконано порівняльний аналіз алгоритмів, обчислювальні результати було порівняно із статистичними даними пасажиропотіку за 2019 рік.

Найважливішим моментом є те, що як і безпосередньо сама теорія графів доволі часто використовується на практиці у реальному житті.

### 3 ОПИС ПРОГРАМНОГО ТА ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ

#### 3.1 Вимоги до програмного забезпечення

Розробка будь-якого програмного забезпечення починається з визначення вимог. Коректно та грамотно сформульовані вимоги є гарантією вдалої та вчасної реалізації програмного продукту на різних етапах розробки. У таблиці 3.1. наведено функціональні вимоги до програмного забезпечення та їх пріоритетність.

Таблиця 3.1 - Функціональні вимоги до програмного забезпечення

№	Вимога	Пріоритет
1	Програма має надавати можливість переглядати список станцій (вершин) графу (метрополітену) та граф метро.	Середній
2	Програма має надавати можливість обчислити класичну центральність, електричну центральність та PageRank.	Високий
3	Програма має надавати можливість побачити порівняльний аналіз методів ранжування у вигляді інтерактивних графіків.	Високий
4	Програма повинна відображати Google Maps зі станціями метро.	Високий
5	Програма має надавати можливість змінити масштаб та пересувати карту	Середній
6	Програма надає можливість зберігати результати	Середній

#### 3.2 Вимоги до технічного забезпечення

Необхідні вимоги до технічного забезпечення наведені у таблиці 3.2.

Таблиця 3.2 - Вимоги до технічного забезпечення

Характеристика	Мінімальне значення
Частота процесора	2 ГГц
Об'єм оперативної пам'яті	6 ГБ
Простір на жорсткому диску	40 ГБ
Операційна система	Windows Server 2013, Windows 10

### 3.3 Засоби розробки

**HTML.** HTML - система форматування для відображення матеріалів, отриманих через Інтернет. Кожна одиниця пошуку відома як веб-сторінка (із Всесвітньої павутини), і такі сторінки часто містять гіпертекстові посилання, що дозволяють отримувати пов'язані сторінки. HTML - це мова розмітки для кодування веб-сторінок. Він був розроблений британським вченим сером Тімом Бернерс-Лі в лабораторії ядерної фізики ЦЕРН у Швейцарії протягом 1980-х років. Теги розмітки HTML визначають такі елементи документа, як заголовки, абзаци та таблиці. Вони розмічають документ для показу комп'ютерною програмою, відомою як веб-браузер. Браузер інтерпретує теги, відображаючи заголовки, абзаци та таблиці в макеті, адаптованому до розміру екрана та доступних йому шрифтів. Документи HTML також містять прив'язки, які є тегами, що визначають посилання на інші веб-сторінки. Якір має вигляд `<a href= "http://www.kpi.com">KPI</a>`, де цитований рядок - це URL-адреса (універсальний локатор ресурсів), на яку вказує посилання (веб-адреса), а наступний за ним текст - це те, що з'являється у веб-браузері, підкреслено, щоб показати, що це посилання на іншу сторінку. Те, що відображається як одна сторінка, також може бути сформовано з декількох URL-адрес, деякі містять текст, а інші графіку.

**CSS.** Основна структура кожної веб-сторінки, HTML, сама по собі дуже проста. Прекрасні веб-сайти, які ви бачите в Інтернеті, стилізовані за допомогою різноманітних інструментів, зокрема CSS.

CSS, або каскадні таблиці стилів, - це мова, яку веб-розробники використовують для стилювання вмісту HTML на веб-сторінці. Якщо цікавить зміна кольорів, типів шрифтів, розмірів шрифтів, тіней, зображень, позиціонування елементів тощо, CSS - це інструмент для роботи.

**JavaScript.** JavaScript є крос-платформною, об'єктно-орієнтованою мовою програмування, яка використовується для створення веб-сторінок інтерактивними (наприклад, із складною анімацією, кнопками, що висувуються, спливаючими меню тощо). Існують також більш вдосконалені версії JavaScript на стороні сервера, такі як Node.js, які дозволяють додати веб-сайту більше функціональних можливостей, ніж просто завантаження файлів (наприклад, співпраця між кількома комп'ютерами в режимі реального часу). Усередині хост-середовища (наприклад, веб-браузера) JavaScript можна підключити до об'єктів свого середовища, щоб забезпечити програмний контроль над ними.

Клієнтський JavaScript розширює основну мову, надаючи об'єктам для управління браузером та його об'єктною моделлю документа (DOM). Наприклад, розширення на стороні клієнта дозволяють програмі розміщувати елементи у формі HTML та реагувати на події користувача, такі як клацання миші, введення форми та навігація сторінкою.

Серверний JavaScript розширює основну мову, надаючи об'єкти, що мають відношення до роботи JavaScript на сервері. Наприклад, розширення на стороні сервера дозволяють програмі взаємодіяти з базою даних, забезпечують безперервність інформації від одного виклику до іншого додатка або виконують маніпуляції з файлами на сервері.

Це означає, що у браузері JavaScript може змінити вигляд веб-сторінки (DOM). І, так само, JavaScript Node.js на сервері може відповідати на власні запити з коду, записаного в браузері.

**DOM.** Об'єктна модель документа (DOM) з'єднує веб-сторінки зі сценаріями або мовами програмування, представляючи в пам'яті структуру документа - наприклад, HTML, що представляє веб-сторінку. Зазвичай це означ

JavaScript, хоча моделювання документів HTML, SVG або XML як об'єктів не є частиною основної мови JavaScript як такої.

DOM представляє документ з логічним деревом. Кожна гілка дерева закінчується вузлом, і кожен вузол містить об'єкти. Методи DOM дозволяють отримати програмний доступ до дерева. За допомогою них можна змінити структуру, стиль або вміст документа.

На вузлах також можуть бути прикріплені обробники подій. Після активації події виконуються обробники подій.

**React.** JavaScript-бібліотека для побудови інтерфейсів, які повинен побачити кінцевий користувач. React використовує віртуальний DOM, тобто, якщо простіше, то це ще одна копія всього HTML-документа в пам'яті. При зміні в ній React вносить зміни в існуючий документ. У React використовується HTML-подібна JSX-розмітка. Це дуже схоже на звичайний HTML, але не є ним. При цьому така розмітка не відображається безпосередньо на сторінці. Спочатку вона перетворюється в той самий віртуальний DOM, а потім вже відображається в потрібному місці сторінки.

**JSON.** JSON (JavaScript Object Notation) - це текстовий формат представлення даних в нотації об'єкта JavaScript. Призначений JSON, також як і деякі інші формати такі як XML і YAML, для обміну даними. JSON в порівнянні з іншими форматами також має досить вагомою перевагою. Він на відміну від них є більш компактним, а це дуже важливо при обміні даними в мережі Інтернет. Крім цього, JSON більш простий у використанні, його набагато простіше читати і писати. При веб-розробці JSON дуже часто застосовується в якості формату для передачі інформації від веб-сервера клієнту (веб-браузеру) при AJAX запиті.

### 3.4 Керівництво користувача

Для того, щоб скористуватись веб-застосунком «Kyiv Underground Centrality» треба написати в google пошуку [kyivcentrality.com.ua](http://kyivcentrality.com.ua). Клієнтська частина системи складається з двох сторінок, на яких зображено google maps із графіками результатів обчислення центральності та форму для введення вхідних

даних. Екранні форми програмного забезпечення наведено у графічному матеріалі.

Для того, щоб передивитись результат обчислення класичної центральності, електричної центральності та PageRank – треба на карті відмітити потрібну станцію метро, одразу ж справа буде відображено графік із порівнянням цих методів. На рисунку 3.1. наведено приклад користування програмою.



Рисунок 3.1 - Приклад користування програмою

### Висновок до розділу

В даному розділі було розглянуто програми, мови програмування, інструменти за допомогою яких було реалізовано систему. Було розглянуто загальні вимоги до технічних засобів. Наведено керівництво користувача для роботи з розробленим програмним забезпеченням.



#### 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

У таблиці 4.1 наведено опис стартап-проекту.

Таблиця 4.1 - Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Розробка та реалізація системи аналізу поведінки людини в веб-середовищі, яка ґрунтується на методах й алгоритмах збору даних з мережі Інтернет.	Для веб-маркетологів	Визначати основні тенденції в інтересах інтернет-користувачів, аналізувати їх на обраних проміжках часу
	Для комерційних представників	Розробляти бізнес-стратегії для просування сайтів та охоплення нової аудиторії
	Для соціологів	Робити висновки про пріоритетність у соціальному житті населення (в т.ч. по регіонах)

Таблиця 4.2 демонструє сильні, слабкі та нейтральні характеристики стартап-проекту.

Таблиця 4.2 - Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№	Техніко-економічні характеристики ідеї	Потенційні концепції конкурентів			W (слабка сторона)	N нейтральна сторона	S сильна сторона
		мій	Google Analytics	Piwik			
1	Візуалізація обсягу трафіку поведінкових показників по днях	+	+	+		+	

Продовження таблиці 4.2

2	Надання класифікації відвідувачів у формі звіту	+	+	-			+
3	Аналіз цінності нового контенту	+	+	+			+

#### 4.1 Технологічний аудит ідеї проекту

Таблиця 4.3 - Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Візуалізувати обсяг трафіку і поведінкові показники по днях	RabbitMQ, PHP	+	+
2	Збір даних про користувача	JavaScript, BeaconAPI	+	+
3	Відображати сегментацію відвідувачів в формі діаграм	PHP, PowerBI	+	+
4	Надавати зміни поведінкових показників		необхідна розробка	+
5	Аналізувати цінність нового контенту	PHP, RabbitMQ	+	+

Висновок: технологічна реалізація проекту є цілком можливою, з поміж названих технологій доцільно використати усі перераховані, адже кожна з них є доступною та наявною на ринку. Технологічний шлях передбачає розробку інструменту для веб-аналітики, який дає змогу відстежити обсяг трафіку і поведінкові показники по днях, а усі отримані дані візуалізувати на клієнтському інтерфейсі.

## 4.2 Аналіз ринкових можливостей запуску стартап-проекту

Таблиця 4.4 - Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку	Характеристика
1	Кількість головних гравців, од	3 (Google Analytics, Piwik, Clicky)
2	Загальний обсяг продаж, грн/ум.од	14 200 грн
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу	Велика конкуренція
5	Специфічні вимоги до стандартизації та сертифікації	_____

Таблиця 4.5 - Характеристика потенційних клієнтів стартап-проекту

Потреба, що формує ринок	Цільова аудиторія	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
Аналіз поведінки людини в веб-середовищі	Веб-маркетологи та власники веб-сайтів	Дотримання плану продажів продукції	Отримання класифікації відвідувачів у формі звіту
	Соціологи	Регіональні або державні соціологічні дослідження	Звітність про пріоритетність у соціальному житті населення
Аналіз цінності нового контенту	Власники компаній та SEO	Унікальність контенту, його сприйнятливості користувачем, кількість відгуків	Спрощення отримання даних для розробки бізнес-стратегії для просування сайтів та охоплення нової аудиторії

Висновок: за результатами аналізу таблиці та попереднім оцінюванням сьогодні ринок є доволі привабливим для входження за умов подолання бар'єру високої конкуренції.

Таблиця 4.6 -Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Конкуренція	Наявність як мінімум трьох серйозних конкурентів на ринку (Google Analytics, Piwik, Clicky)	Порівняння їх основних особливостей, вибір варіанту з більшою кількістю переваг
2	Високий бар'єр входження в технологію	Складність у вивченні технології та всіх пов'язаних з аналітикою особливостей - веде до збільшення фінансових витрат	Відмова від використання технології
3	Час на налаштування всіх функціональних можливостей програмного продукту та потреба у персональному спеціалізованому налаштуванні під конкретні потреби деяких клієнтів	Збільшення часу на розробку та налаштування, а також можливе некоректне відображення даних у перші місяці після запуску	Звернення у службу підтримки продукту

Таблиця 4.7 - Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Аналіз поведінки людини в веб-середовищі	Візуалізувати обсяг трафіку і поведінкові показники по днях	Стрімке визначення необхідного аналітичного інструменту
2	Визначення “портрету” цільового клієнта	Виводити класифікацію відвідувачів у формі звіту	Збільшення доходів компанії та проведення змін в стратегії залучення нових клієнтів
3	Аналіз цінності нового контенту	Одномоментне визначення унікальності контенту, його сприйнятливості користувачем та кількості відгуків	Внесення коректив до бізнес-стратегії для просування сайтів

Таблиця 4.8 - Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства
Тип конкуренції: чиста	На ринку представлено велику кількість продавців і покупців, пропозиція однакових товарів та послуг для продажу	Можливість вільної появи на ринку та виходу з нього покупців та продавців.

Продовження таблиці 4.8

За рівнем конкурентної боротьби: міжнародна	Вид конкурентної боротьби виробників на світовому ринку, яка поєднує внутрішньогалузеву і міжгалузеву конкуренцію.	Така конкуренція сприяє збалансованому розвитку світового ринку, переливу капіталу не тільки між сферами виробництва, а й між різними країнами
За галузевою ознакою: внутрішньогалузева	Це конкуренція між товаровиробниками, що діють в одній галузі народного господарства.	Характеризується зниженням витрат на виробництво й стимулює процес концентрації виробництва та капіталу.
Конкуренція за видами товарів: за бажаннями	Бажання, які споживач може прагнути задовольнити	Розуміння прагнень та актуальних потреб клієнта

Таблиця 4.9 - Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти	Потенційні конкуренти	Постачальники	Клієнти
	Google Analytics, Piwik, Clicky	1.1.1 <u>24Log</u>	Google	Веб-маркетологи, комерційні представники, соціологи
Висновки:	Відомі технологічні рішення, що використовуються для аналізу поведінки людей у веб-	Надає три варіанти лічильника: з детальною статистикою про відвідувачів, лічильник	Сервіс, що надається Google для створення детальної статистики відвідувачів веб-сайтів.	Бажання спрощено отримувати дані для розробки бізнес-стратегії для просування сайтів та

	середовищі	"скільки людей на сайті" і простий лічильник відвідуваності	Статистика збирається на сервері Google, користувач тільки розміщує JS-код на сторінках свого сайту	охоплення нової аудиторії
--	------------	---	---	---------------------------

Висновок: з огляду на конкурентну ситуацію, можна зробити висновок, що загрозу становлять три основні конкуренти: Google Analytics, Piwik, Clicky, а також потенційний конкурент [24Log](#), однак зростаюча динаміка ринку та чиста конкуренція на ньому дають змогу вільно увійти на ринок.

Таблиця 4.10 - Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування
1	Візуалізація обсягу трафіку і поведінкові показники по днях	Аналіз поведінки людини в веб-середовищі
2	Надання класифікації відвідувачів у формі звіту	Визначення "портрету" цільового клієнта
3	Аналіз цінності нового контенту	Одномоментне визначення унікальності контенту, його сприйнятливості користувачем та кількості відгуків
4	Дешевизна програмного продукту	Зацікавленість роботодавців в економії бюджету
5	Зростаюча динаміка ринку	Змога вільно увійти на ринок та мати клієнтів

Таблиця 4.11 - Порівняльний аналіз сильних та слабких сторін.

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів						
			-3	-2	-1	0	+1	+2	+3
1	Візуалізація обсягу трафіку і поведінкові показники по днях	20							+
2	Надання класифікації відвідувачів у формі звіту	20						+	
3	Аналіз цінності нового контенту	15				+			
4	Дешевизна програмного продукту	10					+		
5	Зростаюча динаміка ринку	15							+

Таблиця 4.12 - SWOT- аналіз стартап-проекту

Сильні сторони:	Слабкі сторони:
<ul style="list-style-type: none"> <li>- Візуалізація обсягу трафіку і поведінкові показники по днях;</li> <li>- Надання класифікації відвідувачів у формі звіту;</li> <li>- Аналіз цінності нового контенту;</li> </ul>	<ul style="list-style-type: none"> <li>- Час на налаштування всіх функціональних можливостей програмного продукту;</li> <li>- Можливі помилки у програмі;</li> </ul>
Можливості:	Загрози:
<ul style="list-style-type: none"> <li>- Зростаюча динаміка ринку;</li> <li>- Чистий тип конкуренції в галузі.</li> </ul>	<ul style="list-style-type: none"> <li>- Високий бар'єр входження в технологію;</li> <li>- Наявність як мінімум трьох серйозних конкурентів на ринку (Google Analytics, Piwik, Clicky).</li> </ul>



### 4.3 Розроблення ринкової стратегії проекту

Таблиця 4.14 - Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Веб-маркетологи та SEO	Висока	Високий	Висока	Середня
2	SEO	Висока	Високий	Середня	Середня
3	Соціологи	Середня	Низький	Середня	Середня
Обрано цільову групу: Веб-маркетологи, SEO та SEO.					

Стратегія охоплення ринку: диференційованого маркетингу.

Таблиця 4.15 - Визначення базової стратегії розвитку

Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції	Базова стратегія розвитку
Фланкова атака	Диференційованого маркетингу	Надання класифікації відвідувачів у формі звіту та аналіз цінності нового контенту	Стратегія диференціації

Таблиця 4.16 - Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару ?	Стратегія конкурентної поведінки
Ні, наявність як мінімум трьох серйозних конкурентів на ринку	Намагатиметься залучити нових та привернути увагу клієнтів, що користуються рішенням конкурентів	В основу покладено візуалізацію обсягу трафіку і поведінкові показники користувачів	Стратегія виклику лідера

Таблиця 4.17 - Визначення стратегії позиціонування

Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкуренто-спроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту
Надання класифікації відвідувачів у формі звіту	Стратегія диференціації	візуалізація обсягу трафіку і поведінкові показники по днях; надання класифікації відвідувачів у формі звіту; аналіз цінності нового контенту;	аналізує контент; класифікує відвідувачів дешевий;

#### 4.4 Розроблення маркетингової програми стартап-проекту

Таблиця 4.18 - Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами
1	Визначення “портрету” цільового клієнта	Виводити класифікацію відвідувачів у формі звіту	Застосування тільки сучасних алгоритмів та усунення невдалих моделей, застосованих конкурентами
2	Аналіз цінності нового контенту	Одномоментне визначення унікальності контенту, його сприйнятливості користувачем та кількості відгуків	Можлива інтеграція з існуючими спеціалізованими системами для моніторингу якості контенту

Продовження таблиці 4.19

3	Зацікавленість SEO в економії бюджету	Дешевизна програмного продукту	Можливість безкоштовної установки

Таблиця 4.20 - Визначення меж встановлення ціни

Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар
0 - 700 \$	0 - 150 000 \$	300 - 5 000 \$	0 - 550 \$

Обрано стратегію диференційованого маркетингу для охоплення ринку, що орієнтована на веб-маркетологів, SEO та SEO.

### Висновки до розділу

Даний стартап проект може аналізувати поведінки людини в веб-середовищі, яка ґрунтується на методах й алгоритмах збору даних з мережі Інтернет. Даний проект є новим в своїй галузі, що дає непогані шанси на успіх.

## ВИСНОВКИ

В дисертації було досліджено методи аналізу мереж різного призначення. Було розглянуто методи та моделі аналізу мереж різного призначення. Описано структуру бізнес процесів у вигляді контекстної діаграми та наведено схеми структурні варіантів використання. Були визначені цілі та завдання магістерської дисертації. Було розроблено рішення з інформаційного забезпечення системи.

Наведено змістовну та математичну постановку задачі визначення центральності вершин (станцій) київського метрополітену. Проведено огляд існуючих методів розв'язання задачі, результати експериментів і порівняння результатів алгоритмів представлено представлено у вигляді таблиці та графіків. Було обчислено та порівняно класичну центральність, електричну центральність та PageRank-центральність.

Було проведено серію експериментів для обчислення класичної центральності, електричної центральності та PageRank-центральності для усіх вершин (станцій) графу (київського метрополітену). Результати експериментів показали, що найбільш завантаженою, за підрахунками кількості поїздок, виявилася станція «Лісова». Також серед найпопулярніших – «Академмістечко» та «Вокзальна». Натомість станцією «Дніпро» користуються в 20 разів рідше. Загалом найбільший пасажиропотік – на червоній гілці метрополітену.

Було розроблено стартап проект, який може аналізувати поведінку людини в веб-середовищі, яка ґрунтується на методах й алгоритмах збору даних з мережі Інтернет. Даний проект є новим в своїй галузі, що дає непогані шанси на успіх.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Батура Т.В. Методы анализа компьютерных социальных сетей / Т.В. Батура // Вестник Новосибирского государственного университета. Серия: Информационные технологии. – 2012. – Т. 10, № 4. – С. 13–28.
2. Коломейченко М.И. Алгоритм выделения сообществ в социальных сетях / М.И. Коломейченко, А.А. Чеповский, А.М. Чеповский // Фундаментальная и прикладная математика. – 2014. – Т. 19, № 1. – С. 21–32.
3. E. Otte, R. Rousseau, Social network analysis: a powerful strategy, also for the information sciences, J. Inf. Sci. 28 (6) (2002) 441–453
4. I. King, J. Li, K.T. Chan, A brief survey of computational approaches in social computing, in: 2009 International Joint Conference on Neural Networks, IEEE, pp. 1625-1632.
5. G. Barbier, C. Liu H. Data Mining in Social Media. In: Aggarwal, Data mining in social media, in: C. Aggarwal (Ed.), Social Network Data Analytics, Springer, Boston, MA, USA, 2011, [http://dx.doi.org/10.1007/978-1-4419-8462-3\\_12](http://dx.doi.org/10.1007/978-1-4419-8462-3_12).
6. C.C. Aggarwal, H. Wang, Text mining in social networks, in: C.C. Aggarwal (Ed.), Social Network Data Analytics, Springer US, Boston, MA, USA, 2011, [http://dx.doi.org/10.1007/978-1-4419-8462-3\\_13](http://dx.doi.org/10.1007/978-1-4419-8462-3_13).
7. P. Gundecha, H. Liu, Mining social media: a brief introduction, in: New Directions in Informatics, Optimization, Logistics, and Production, 2012, pp. 1–17, Informs..
8. M. Adedoyin-Olowe, M.M. Gaber, F. Stahl, A survey of data mining techniques for social media analysis. arXiv preprint arXiv:1312.4617 2013.
9. G. Nandi, A. Das, A survey on using data mining techniques for online social network analysis, Int. J. Comput. Sci. Issues (IJCSI) 10 (6) (2013) 162.
10. A. Guille, H. Hacid, C. Favre, D.A. Zighed, Information diffusion in online social networks: A survey, ACM Sigmod Rec. 42 (2) (2013) 17–28

11. I. Guellil, K. Boukhalfa, Social big data mining: A survey focused on opinion mining and sentiments analysis, in: 2015 12th International Symposium on Programming and Systems (ISPS), IEEE, 2015, pp. 1–10.
12. K. Aggarwal, K. Kapoor, J. Srivastava, Data mining techniques for social networks analysis, in: R. Alhajj, J. Rokne (Eds.), Encyclopedia of Social Network Analysis and Mining, Springer New York, New York, NY, USA, 2017, [http://dx.doi.org/10.1007/978-1-4614-7163-9\\_56-1](http://dx.doi.org/10.1007/978-1-4614-7163-9_56-1).
13. John Hopcroft and Daniel Sheldon. Manipulation-resistant reputations using hitting time. Technical report, Cornell University, 2007.
14. Hanneman, Robert A. and Mark Riddle. Introduction to social network methods. – Riverside, CA: University of California, Riverside, 2005. – 322 p. (published in digital form at <http://faculty.ucr.edu/~hanneman/> )
15. Бойко Е. Кластеризация социальных сетей с помощью алгоритма кластеризации BSP / Е. Бойко // Восточно-Европейский журнал передовых технологий. – 2012. – Т. 3, № 11 (57). – С. 34–36.
16. Бойко Е. Кластеризация социальных сетей с помощью алгоритма кластеризации BSP / Е. Бойко // Восточно-Европейский журнал передовых технологий. – 2012. – Т. 3, № 11 (57). – С. 34–36.
17. Губанов Д. Нечеткие модели влияния в социальных сетях / Д. Губанов, Д. Новиков, А. Чхартишвили // Труды VI Всероссийской школы-семинара «Управление большими системами». – 2009. – Т. 2. – С. 141–145.
18. Мелешко Є., Гермак В., Якименко М. Дослідження методів аналізу соціальних мереж з точки зору інформаційної безпеки держави, Системи управління, навігації та зв'язку, Кіровоград, Україна, 2015.
19. Губанов Д.А. Социальные сети: модели информационного влияния, управления и противоборства / Д.А. Губанов, Д.А. Новиков, А.Г. Чхартишвили. – М.: Физматлит, 2010.
20. Гнидко К.О. Контроль потенциально опасного информационно-психологического воздействия на индивидуальное и групповое сознание

потребителей мультимедийного контента / К.О. Гнидко, А.Г. Ломако // Труды СПИИРАН. – 2015. – Вып. 1(38). – С. 9-33.

21. T. Haveliwala. Topic-sensitive pagerank. In Proceedings of the Eleventh International World Wide Web Conference, 2002.
22. David Aldous and James Allen Fill. Reversible Markov Chains. 2001.
23. L. Katz. A new status index derived from sociometric analysis. In Psychometrika, 1953.
24. X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In ICML, volume 20, 2003.
25. A. Balmin, V. Hristidis, and Y. Papakonstantinou. ObjectRank: Authority-based keyword search in databases. In VLDB, 2004.
26. Soumen Chakrabarti. Dynamic personalized pagerank in entity-relation graphs. In Proc. WWW, New York, NY, USA, 2007.
27. Qiaozhu Mei, Dengyong Zhou, and Kenneth Church. Query suggestion using hitting time. In CIKM '08, pages 469–478, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-991-3
28. M. Brand. A Random Walks Perspective on Maximizing Satisfaction and Profit. In SIAM '05, 2005.
29. M. Saerens, F. Fouss, L. Yen, and P. Dupont. The principal component analysis of a graph and its relationships to spectral clustering, 2004.
30. Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. ACM Transactions on Graphics, 23:689–694, 2004.
31. Leo Grady. Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(11):1768–1783, 2006. ISSN 0162-8828.
32. X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In ICML, volume 20, 2003.
33. G. Jeh and J. Widom. Scaling personalized web search. In Stanford University Technical Report, 2002.

34. D. Fogaras, B. Rcz, K. Csalogány, and Tamás Sarlós. Towards scaling fully personalized pagerank: Algorithms, lower bounds, and experiments. *Internet Mathematics*, 2004.
35. D. Spielman and N. Srivastava. Graph sparsification by effective resistances. In *Proceedings of the STOC'08*, 2008.
36. P. Berkhin. Bookmark-Coloring Algorithm for Personalized PageRank Computing. *Internet Mathematics*, 2006.
37. Tamas Sarlos, Andras A. Benczur, Karoly Csalogany, Daniel Fogaras, and Balazs Racz. To randomize or not to randomize: space optimal summaries for hyperlink analysis. In *WWW*, 2006.
38. Purnamrita Sarkar and Andrew Moore. A tractable approach to finding closest truncated-commute-time neighbors in large graphs. In *Proc. UAI*, 2007.
39. Langville A.N., Meyer C.D. *Google's PageRank and beyond: The science of search engine rankings*. – Princeton University Press, 2015.
40. L. Backstrom and J. Leskovec. Supervised random walks: predicting and recommending links in social networks. In *WSDM*, 2011
41. C. Faloutsos, J. Han, and P. S. Yu., editors. *Link Mining: Models, Algorithms and Applications*. 2010.
42. Stephen J. Hardiman and Liran Katzir. 2013. Estimating clustering coefficients and size of social networks via random walk. In *Proceedings of WWW 2013*. 539--550.
43. Lai D, Lu H, Nardini C (2010) Finding communities in directed networks by pagerank random walk induced network embedding. *Physica A Stat Mech Appl* 389:2443–2454
44. J. Zhao, P. Wang, J. C. S. Lui, D. Towsley and X. Guan, Sampling online social networks by random walk with indirect jumps, 2017, [online] Available: <https://www.arxiv.org/abs/1708.09081>.
45. Котов Э.М., Методы анализа гиперссылок при информационном поиске в глобальной сети // *Известия южного федерального университета. Технические науки* – 2012, с. 233- 237



46. A. Nazi, Z. Zhou, S. Thirumuruganathan, N. Zhang, and G. Das. Walk, not wait: Faster sampling over online social networks. *PVLDB*, 8(6):678--689, 2015.
47. S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: joint friendship and interest propagation in social networks. In *WWW*, 2011.
48. Huaijun Qiu and Edwin R. Hancock. Image segmentation using commute times. In *Proc. BMVC*, 2005.
49. T. Haveliwala. Topic-sensitive pagerank. In *Proceedings of the Eleventh International World Wide Web Conference*, 2002.
50. Zoltan Gyongyi, Hector Garcia-Molina, and Jan Pedersen. Combating web spam with trustrank. In *VLDB'2004: Proceedings of the Thirtieth international conference on Very large data bases*. VLDB Endowment, 2004.
51. Baoning Wu and Kumar Chellapilla. Extracting link spam using biased random walks from spam seed sets. In *WWW*, 2007.
52. L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt. Shape representation and classification using the poisson equation. In *CVPR*, 2004.
53. John Lafferty and Chengxiang Zhai. Document language models, query models, and risk minimization for information retrieval. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001.
54. Einat Minkov, William W. Cohen, and Andrew Y. Ng. Contextual search and name disambiguation in email using graphs. In *SIGIR '06*, 2006.
55. Kristina Toutanova, Christopher D. Manning, and Andrew Y. Ng. Learning random walk models for inducing word dependency distributions. In *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, 2004.
56. Francois Fouss, Alain Pirotte, Jean michel Renders, and Marco Saerens. A novel way of computing dissimilarities between nodes of a graph, with

- application to collaborative filtering. In ECML workshop on Statistical Approaches for Web Mining, 2004.
57. Amruta Joshi, Ravi Kumar, Benjamin Reed, and Andrew Tomkins. Anchor-based proximity measures. In WWW '07: Proceedings of the 16th international conference on World Wide Web, 2007.
  58. A. Smola and R. Kondor. Kernels and regularization on graphs. In Springer Verlag, editor, Learning Theory and Kernel Machines, 2003.
  59. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In Proc. WWW, 1998.
  60. Sepandar D. Kamvar, Taher H. Haveliwala, Christopher D. Manning, and Gene H. Golub. Extrapolation methods for accelerating pagerank computations. In WWW '03: Proceedings of the 12th international conference on World Wide Web, 2003.
  61. David Liben-Nowell and Jon Kleinberg. The link prediction problem for social networks. In CIKM '03, 2003
  62. Chebukov D., Izaak A., Misurina O., Pupyrev Yu., Zhizhchenko A. Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today. Lecture Notes in Computer Science. 2013. Vol. 7961. P. 344–348.
  63. Zhizhchenko A.B., Izaak A.D. Informatsionnaya sistema Math-Net.Ru. *Primenenie sovremennykh tekhnologii v nauchnoi rabote matematika* [Informational system Math-Net.Ru. Application of modern technologies in mathematical scientific work]. *Uspekhi Matematicheskikh Nauk* [Russian Mathematical Surveys]. 2007. Vol. 62, N 5 (377). P. 107–132.
  64. Brin S. and Page L. The anatomy of a largescale hypertextual Web search engine. *Computer Networks and ISDN Systems*. 1998. Vol. 30, N 17. P. 107–117.
  65. Jeong, H., Mason, S., Barab'asi, A.L., Oltvai, Z.: Lethality and centrality in protein networks. *Nature* 411, 41–42 (2001)

66. Coffman, T., Greenblatt, S., Marcus, S.: Graph-based technologies for intelligence analysis. *Communications of the ACM* 47(3), 45–47 (2004)
67. Myerson R. B. Graphs and cooperation in games. *Math. Oper. Res.* 1977. Vol. 2. P. 225–229.
68. Mazalov V. V., Trukhina L. I. Generating functions and the Myerson vector in communication networks. *Discrete Mathematics and Applications*. 2014. Vol. 24, N 5. P. 295–303.
69. Brandes U., Fleischer D. Centrality measures based on current flow. In *Proceedings of the 22nd annual conference on Theoretical Aspects of Computer Science*. 2005. P. 533–544.
70. Newman M. E. J. A measure of betweenness centrality based on random walks. *Social networks*. 2005. Vol. 27. P. 39–54.
71. Avrachenkov K., Litvak N., Medyanikov V., Sokol M. Alpha current flow betweenness centrality. *Lecture Notes in Computer Science*. 2013. Vol. 8305. P. 106–117.
72. Mazalov V.V., Trukhina L.I. Generating functions and the Myerson vector in communication networks // *Discrete Mathematics and Applications*. Vol. 24(5). 2014. Pp. 295-303.
73. Вихляєва А.О. Аналіз соціальних мереж/ В.Д. Попенко, А.В. Вихляєва// Матеріали IX Міжнародної науково-практичної конференції «DYNAMICS OF THE DEVELOPMENT OF WORLD SCIENCE» - м. Осака, 17-19 травня 2020 р. – С. 488-504.
74. Вихляєва А.О. «Аналіз моделей ранжування вершин у графах мереж різного призначення». / А.О. Вихляєва, В.Д. Попенко // Матеріали V всеукраїнської науково-практичної конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2020) – м. Київ.: НТУУ «КПІ ім. Ігоря Сікорського», 26-27 листопада 2020 р. – С. 52-58.

## **ДОДАТОК А Графічний матеріал**

## Схема структурна варіантів використання

### СХЕМА СТРУКТУРНА ВАРІАНТІВ ВИКОРИСТАННЯ



Демонстраційний плакат до магістерської дисертації

*“Аналіз моделей ранжування вершин у графах мереж різного призначення”*

Виконала студентка гр. ІС-92мп

Вихляєва А.О.

Керівник

Попенко В.Д.

Схема структурна контекстної моделі системи

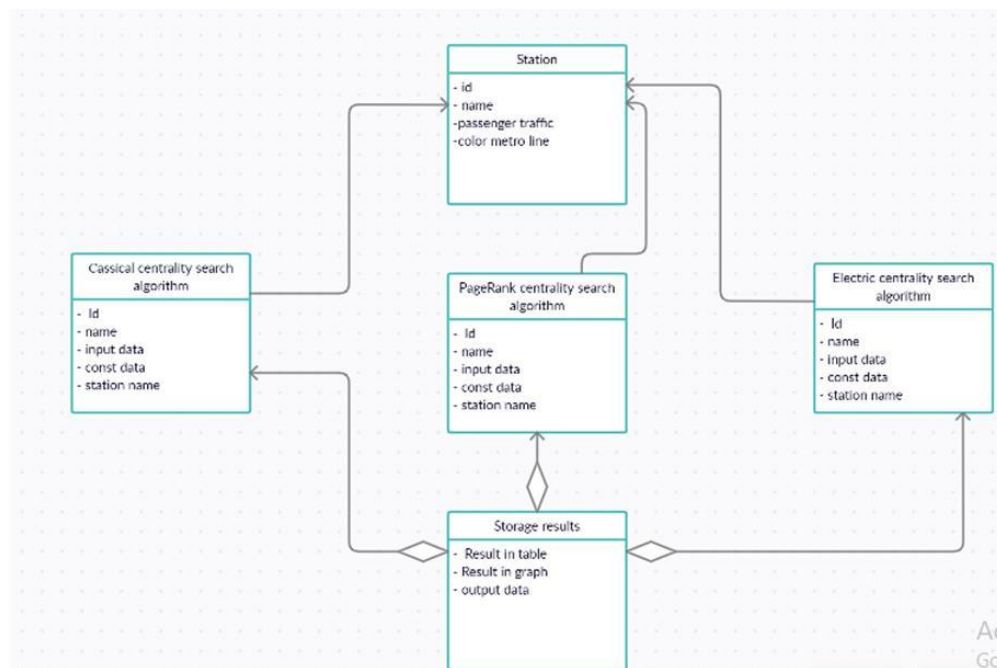
СХЕМА СТРУКТУРНА КОНТЕКСТНОЇ МОДЕЛІ



Демонстраційний плакат до магістерської дисертації  
"Аналіз моделей ранжування вершин у графах мереж різного призначення"  
Виконала студентка гр. ІС-92мп  
Керівник  
Вихляєва А.О.  
Попенко В.Д.

## Схема структурна класів програмного забезпечення

### СХЕМА СТРУКТУРНА КЛАСІВ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ



Демонстраційний плакат до магістерської  
дисертації  
"Аналіз моделей ранжування вершин у графах мереж різного  
призначення"

Виконала студентка гр. ІС-92мп

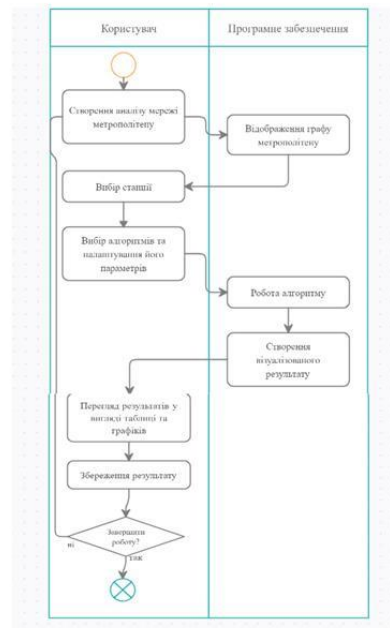
Вихляєва А.О.

Керівник

Попенко В.Д.

## Діаграма діяльності роботи програми

### ДІАГРАМА ДІЯЛЬНОСТІ РОБОТИ ПРОГРАМИ



Демонстраційний плакат до магістерської  
дисертації  
“Аналіз моделей ранжування вершин у графах мереж різного  
призначення”

Виконала студентка гр. ІС-92мп

Вихляєва А. О.

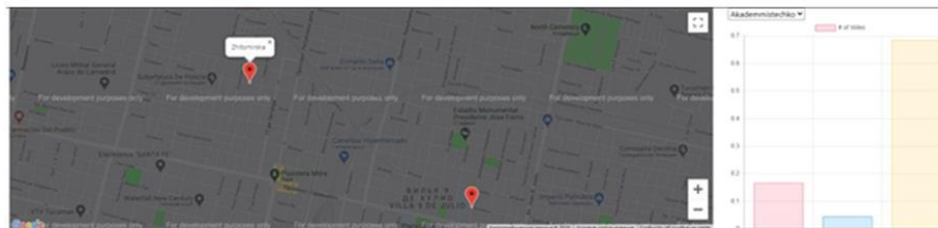
Керівник

Попенко В.Д.



## Екранні форми

### ЕКРАННІ ФОРМИ



Демонстраційний плакат до магістерської  
дисертації  
“Аналіз моделей ранжування вершин у графах мереж різного  
призначення”

Виконала студентка гр. ІС-92мп

Вихляєва А.О.

Керівник

Попенко В.Д.

Попенко В.Д.